BIAS REDUCTION IN NONPARAMETRIC HAZARD RATE ESTIMATION

by

BAGKAVOS IOANNIS DIMITRIOS

A thesis submitted to the Faculty of Science
of The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Mathematics and Statistics
Faculty of Science
The University of Birmingham

19 September 2002 .

CONTENTS

1 Introduction **1**
1.1 General about the failure rate. 1
1.2 Kernel based estimators. 3
1.3 Mean Square Error analysis. 6
1.4 Bandwidth selection. 8
1.4.1 Direct plug-in rules. 9
1.4.2 Least squares cross validation. 10
1.5 Problems with kernel-based hazard rate estimators. 10
1.5.1 Boundary bias. 11
1.5.2 Bias reduction. 12
1.5.3 Transformations. 13
1.6 Outline of the thesis. 14

2 Hazard Rate Estimation using Local Linear Fit **15**
2.1 Introduction. 15
2.2 An empirical estimator of the hazard rate. 16
2.3 Local linear fitting. 20
2.4 Asymptotic properties. 21
2.5 Bandwidth selection. 25
2.5.1 Simple plug-in rule. 26
2.5.2 Least squares cross validation. 31

3 Variable Bandwidth For Hazard Rate Estimation **33**
3.1 Introduction. 33
3.2 Variable bandwidth estimators. 34
3.3 Mean square error analysis of the ideal variable bandwidth estimators. . 35
3.3.1 Mean and variance of the $\tilde{\lambda}_{n,1}$ estimator. 35
3.3.2 Mean and variance of the $\tilde{\lambda}_{n,2}$ estimator. 36
3.4 Adaptive estimators. 44
3.5 Comparison of the ideal and adaptive estimators. 44
3.5.1 Comparison of the ordered estimators. 44
3.5.2 Comparison of the ratio estimators. 47
3.6 Proof of theorems 3.5.1 and 3.5.2. 48
3.6.1 Proof of theorem 3.5.1. 48
3.6.2 Proof of theorem 3.5.2. 56
3.7 Lemmas. 62

4 Transformations in failure rate estimation **69**
4.1 Introduction. 69
4.2 Motivation. 69
4.3 Asymptotic error of the estimator. 71
4.3.1 Lemmas. 76
4.4 Implementation of $\hat{\lambda}_{n,2}(x|h_1, h_2)$ and $\hat{\lambda}_{T,2}(x|h_1, h_2)$. 83
4.4.1 Interactive bandwidth choice. 83
4.4.2 Implementation of the estimators. 84

5 Conclusions and suggestions for further research **88**
5.1 Concluding remarks. 88
5.2 Ideas for future work. 89

A Calculations **90**
A.1 Calculations, chapter 2. 90
A.1.1 Minimization - local linear fit. 90
A.1.2 Proof of lemma 2.4.1. 91
A.1.3 Functionals of the hazard rate. 94
A.1.4 Estimation of ψ_r with reference to $W(\beta, \kappa)$. 95
A.1.5 Mean and variance of $\hat{\psi}_r(g)$. 95
A.2 Calculations, chapter 3. 101
A.2.1 Multiplication. 101

LIST OF TABLES

1	**Introduction**	**1**
2	**Hazard Rate Estimation using Local Linear Fit**	**15**
3	**Variable Bandwidth For Hazard Rate Estimation**	**33**
4	**Transformations in failure rate estimation**	**69**
5	**Conclusions and suggestions for further research**	**88**
A	**Calculations**	**90**

LIST OF FIGURES

1 Introduction **1**

1.1 Estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ 5

2 Hazard Rate Estimation using Local Linear Fit **15**

2.1 Local linear fit estimator using the plug-in rule bandwidth 30

3 Variable Bandwidth For Hazard Rate Estimation **33**

4 Transformations in failure rate estimation **69**

4.1 Dialog windows for interactive bandwidth choice 84
4.2 Transformed, and adaptive estimators estimating the hazard rate of χ^2_{12} distribution. 85
4.3 Transformed and adaptive estimators estimating the hazard rate of the $LN(0,1)$ distribution. 86
4.4 Transformed and adaptive estimators estimating the hazard rate of the normal mixture $\frac{2}{3}N(2,1)+\frac{1}{3}N(3,.2^2)$. 86
4.5 Estimation of the hazard rate by the transformed (green line) and the adaptive (blue line) estimates for the suicide data. 87

5 Conclusions and suggestions for further research **88**

A Calculations **90**

SYNOPSIS

The need of improvement of the bias rate of convergence of traditional nonparametric hazard rate estimators has been widely discussed in the literature. Initiated by recent developments in kernel density estimation we distinguish and extend three popular bias reduction methods to the hazard rate case.

A usual problem of fixed kernel hazard rate estimates is their poor performance at endpoints. Noticing the automatic boundary adaptive property of the local linear smoother (Fan and Gijbels [13]) we adapt the method to the hazard rate case and we show that it results in estimators with bias at endpoints reduced to the level of interior bias. We then turn our attention to global bias problems. Utilizing the proposals of Hall and Marron [16] for estimation using location varying bandwidth as a means to improve the bias rate of convergence, we extend two distinct hazard rate estimators to the point that they make use of the method. The theoretical study of the resulting estimators verifies this improvement. A somewhat related way of improvement over the ordinary kernel estimates of the hazard rate is attained by extending the method of empirical transformations (Ruppert and Cline [35]). Studying the asymptotic square error of the resulting estimator we show that the advance is similar to the variable bandwidth approach.

In summarizing the thesis, ideas and plans for further work are suggested.

ACKNOWLEDGEMENTS

I would like to deeply thank my supervisor Dr. P.N. Patil and Dr R.A. Atkinson for providing me with the topic and for their patience and help during the work undertaken herein. I would also like to thank my family and all my friends for their encouragement.

I am indebted to the Institute of State Scholarships (IKY) of Greece for their help and financial support throughout my time at The University of Birmingham.

Chapter 1

INTRODUCTION

1.1 General about the failure rate.

In survival analysis we are interested in the probability that an 'item' functioning at time x will stop 'functioning', at most once, in the interval $(x, x + dx)$. The terms 'item' and 'functioning' here are used in their broader sense. Typical examples of this situation can be a component of a device or a patient under treatment. Mathematically we write this probability as

$$\lambda(x) = \lim_{dx \to 0+} \frac{\mathrm{P}(x \leq X < x + dx | x \leq X)}{dx}$$

where the random variable $X > 0$ denotes the lifetime of the article we are interested in. The function $\lambda(x)$ is called the *hazard* or *failure rate* function. By the definition of conditional probability we write this limit as

$$\lambda(x) = \frac{f(x)}{1 - F(x)}, \quad F(x) < 1$$

where $f(x)$ and $F(x)$ are respectively the probability density function (pdf) and the cumulative distribution function (cdf) of X. Then we can define the integrated hazard rate or cumulative hazard function of the distribution as

$$\Lambda(x) = \int_0^x \lambda(t)\, dt.$$

The hazard rate and the cumulative hazard function provide two equivalent mathematical ways for studying the failure time, X, of an item. To see that, integrating $\lambda(x)$ we get

$$\Lambda(x) = \int_0^x \frac{f(t)}{1 - F(t)}\, dt = -\log\big[1 - F(x)\big].$$

Equivalently,

$$1 - F(x) = e^{-\Lambda(x)}$$

and thus,

$$f(x) = \lambda(x) e^{-\Lambda(x)}.$$

In the present thesis we confine ourselves to the study of univariate failure rate as defined above. For multivariate representation of the failure rate, see chapter 10 in [9] and paragraph 4.3 in [33].

There are many examples of the use of the hazard rate function in real life situations. To name but a few, the hazard rate can be used to describe the failure times of manufactured items which are subjected to life tests in order to obtain information on their endurance. Some types of manufactured items can be repaired should they fail. In this case the hazard rate is used to describe the time between two successive failures of an item. In a totaly different setting, the hazard rate is used to describe the survival times of individuals with a certain disease started from the day of diagnosis or some other start point.

There are many reasons to introduce the hazard rate or the cumulative hazard function when a pdf or a cdf of a random variable (r.v.) X can give equivalent information. For example one can attribute the popularity of the idea of hazard to the natural desire to reckon things 'as of now'. That is, it may be physically informative to consider the immediate 'risk' attached to an individual known to be alive at time x. Further, hazard models are particularly convenient to work with when we have censoring of observations, and this is very usual in areas such as reliability and medical studies. In reliability studies, comparison of the hazard rates of two or more types of components is more informative than comparison of their pdf's. Finally a rather special property of the hazard rate is that the hazard function of time related events is often time varying in a structured fashion which can sometimes be decomposed to additive phases, where each of them is in general shaped by a different generic hazard function. This is a very convenient representation of a model because it gives us the freedom to adopt a different hazard function for each case. To see that mathematically consider an array of modes of failure $M_1, \ldots, M_n$ (failure in any one means failure of the entity and types of failure are statistical independent) with individual hazards $\lambda_1(t), \lambda_2(t), \ldots, \lambda_n(t)$ then the corresponding cumulative hazard rate function is $\sum_i \lambda_i(t)$. Particularly in this case it may be of interest to use a method that offers wide flexibility in the possible form of the generic function modelling each phase.

Thus estimation of hazard rate consists an important problem in areas such as reliability and survival analysis. The classical method of estimation involves modelling the failure times by appropriate failure time probability distributions such as exponential, gamma, weibull, etc. This results in a functional form for the hazard rate which is known up to a parameter or a vector of parameters. Hence estimation of the hazard rate reduces to estimation of these parameters. There is quite a rich literature on parametric inference of hazard rate estimation see for example Barlow and Proschan [3], Kalbfleisch and Prentice [20] and Lawless [21]. Though parametric methods for hazard rate estimation are very useful and efficient and thus have the advantage of great accuracy, the inference is meaningful only if the assumed model is at least approximately true. This leads to alternative approaches based in nonparametric methods. These methods make no formal assumptions about the mechanism that generates the sample other than that it is a random sample, thus they offer more flexibility. For a detailed account of nonparametric methods see Prakasa Rao [33] or Simonoff [40].

In the next section we give a brief introduction to nonparametric kernel based methods for density estimation and thereby for hazard rate estimation.

1.2 Kernel based estimators.

Kernel estimation of ratio functions such as the hazard rate is usually done by either estimating separately the numerator and the denominator or by treating the hazard rate as a function by itself, see for example Patil *et al.* [30] and the references therein. In this section we show how this is done and we demonstrate the performance of the resulting estimators.

Suppose that we have a sample $X_1, X_2, \ldots, X_n$ from a probability density f and distribution function F. From the definition of the pdf we have

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P(x - h < X < x + h).$$

A natural estimator of $f(x)$ then is

$$\hat{f}(x) = \frac{1}{2nh} \{\# X_i \in (x - h, x + h)\}$$

for any given small h. This is the naive estimator of the unknown density function, and we can write it mathematically as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where,

$$K(x) = \begin{cases} 1/2 & |x| < 1 \\ 0 & \text{otherwise}. \end{cases}$$

One can improve this estimator by using a more general form for the function K, the *kernel*. Generally it is required that the kernel be a symmetric function, integrating to one, with the additional properties that

$$\int |K(x)|\, dx < +\infty, \quad \lim_{x \to +\infty} |xK(x)| = 0.$$

Thus, the idea behind the kernel estimate is that the unknown density is being estimated as an average of densities, each one of them centered at each observation. Using the shorthand notation $K_h(\cdot) = h^{-1} K(\cdot/h)$, we can rewrite the above formula as

$$\hat{f}(x|h) = \frac{1}{n} \sum_{i=1}^{n} K_h\left(x - X_i\right).$$

The cdf can then be estimated either by integrating the above function or by using the empirical cumulative distribution function (ecdf). Let

$$\hat{F}_1(x) = \int_0^x \hat{f}(u)\, du \quad \text{and,}$$

$$\hat{F}_2(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \le x\}},$$

where I is an indicator function. Then a kernel estimate of the hazard rate is

$$\hat{\lambda}(x) = \frac{\hat{f}(x)}{1 - \hat{F}(x)}$$

where $\hat{F}(x)$ is either $\hat{F}_1$ or $\hat{F}_2$. Note that for $\hat{F} = \hat{F}_2$, $\hat{\lambda}(x)$ is not well defined for $x > X_{(n)}$, where $X_{(n)}$ denotes the nth order statistic. For this reason we use a modified version of the $\hat{F}_2$ estimator given by

$$F_n(x) = \frac{\{\# X_i \le x\} - 1}{n} = \hat{F}_2(x) - \frac{1}{n}.$$

Note that for large n, $\hat{F}_2$ and F_n are equivalent. Thus taking $\hat{F} = F_n$ we define $\lambda_1(x)$, one of the kernel estimators of the hazard rate discussed in this dissertation, as

$$\hat{\lambda}_1(x|h) = \frac{\hat{f}(x|h)}{1 - F_n(x)}. \tag{1.1}$$

Now, as opposed to ratio of two functions, consider the hazard rate function, $\lambda(x)$, as a function by itself. Using a kernel function, as defined earlier, on a small compact interval, $\lambda(x)$ can be approximated as

$$\begin{aligned}
\lambda(x) \simeq \lambda^*(x) &= \int K_h(x-u)\lambda(u)\,du \\
&= \int K_h(x-u)\,d\Lambda(u) \\
&= \int K_h(x-u)\,d\Lambda_n(u) + \int K_h(x-u)\,d\left[\Lambda(u) - \Lambda_n(u)\right],
\end{aligned} \tag{1.2}$$

where $\Lambda_n(x)$ is an empirical version of $\Lambda(x)$ defined as

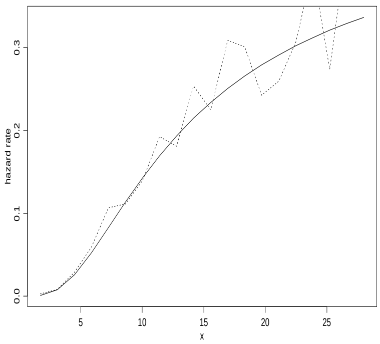
$$\Lambda_n(x) = -\log\left(1 - F_n(x)\right).$$

Now, the first term on the right hand side (RHS) of (1.2) provides us with an estimator of λ. Thus we write

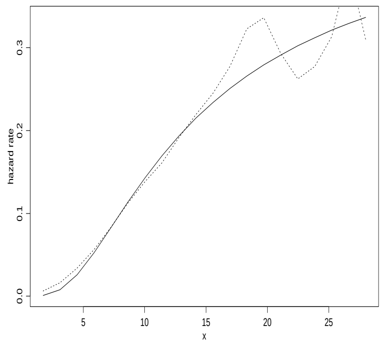
$$\hat{\lambda}_2(x|h) = \int K_h(x-u)\,d\Lambda_n(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{K_h\left(x - X_i\right)}{1 - F_n(X_i)} = \sum_{i=1}^{n}\frac{K_h\left(x - X_{(i)}\right)}{n - i + 1}. \tag{1.3}$$

It may be noted that the approximation error $\lambda^*(x) - \lambda(x)$ represents the bias in the estimation procedure whereas the second term in (1.2) represents the noise. Also from (1.2) it can be seen that small values of h will lead to small values of bias and large noise and large values of h will lead to large bias and small noise. This is demonstrated graphically in the next paragraph and an exact quantification of these quantities and their dependencies on h is given in section 1.4.

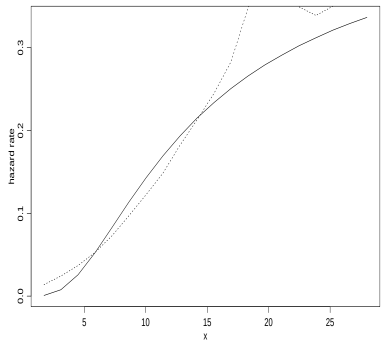
Essentially h determines the spread of the kernel, in other words it controls the amount of smoothing applied to an estimator. For this reason it is called the smoothing parameter or *bandwidth*. In figure 1.1 we use a sample of 800 values from the χ^2_{12} distribution to estimate the true hazard rate for different bandwidth values so that we



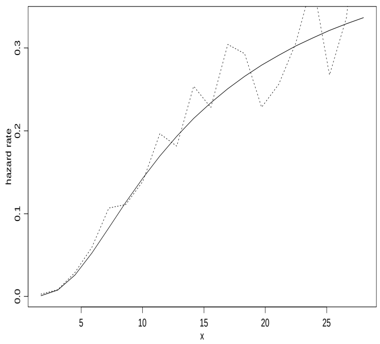
(a) $h = 0.3$



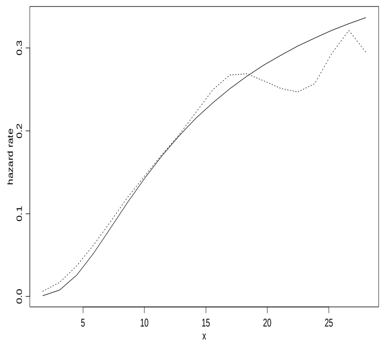
(b) $h = 1.8$



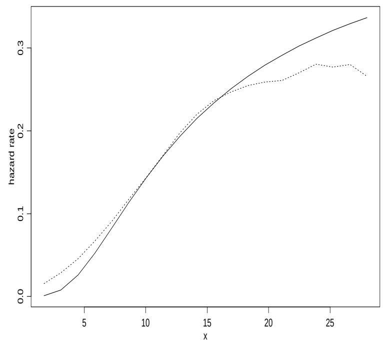
(c) $h = 3$



(d) $h = 0.3$



(e) $h = 1.8$



(f) $h = 3$

Figure 1.1: Top row: estimator $\hat{\lambda}_1$ (dashed line). Second row: estimator $\hat{\lambda}_2$ (dashed line). In all figures the true hazard function is the solid line.

can see how it affects the performance of an estimator. The choice of this particular distribution for the demonstration is justified by the fact that it also shows the effect of the denominator of the estimator on the variance.

First, note the poor performance of the estimators from $x = 15$ onwards. It is due to lack of enough data beyond that point and, as it will be noted later, to an increase in the variance of the estimator for larger x. For this reason we choose to demonstrate the performance of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ up to $x = 28$. In figures 1.1(a) and 1.1(b) one can see that with small h the estimators are unstable, having large variance throughout the range of x. In figures 1.1(b) and 1.1(e) this particular choice of bandwidth has reduced the variance, keeping at the same time the bias at reasonably low levels. Even higher values of h, as in figures 1.1(c) and 1.1(f), result to very smooth estimates with the variance very much reduced but at the expense of larger bias. This is apparent from the behavior of the estimators up to $x = 15$. These figures suggest that large bandwidth values lead to oversmoothing and as a result hide features of the real curve whereas small values lead to undersmoothing and thus give the impression that the curve has features which are not actually there.

1.3 Mean Square Error analysis.

Typically the accuracy of an estimator $\hat{\lambda}$ of a hazard rate λ is measured by suitable error criteria. If the objective is to assess the performance at a point then the most commonly used error criteria are the squared error, $(\hat{\lambda}-\lambda)^2$ and the absolute deviation, $|\hat{\lambda}-\lambda|$. If, however, the objective is to assess the performance of an estimator $\hat{\lambda}$ as a whole curve then this is done by using either the L_1, the L_p or the L_∞ distances which are defined by

$$\begin{aligned} L_1(\lambda) &= \int \left|\hat{\lambda}-\lambda\right|, \\ L_p(\lambda) &= \left(\int \left|\hat{\lambda}-\lambda\right|^p\right)^{\frac{1}{p}}, \quad 2 \le p < +\infty \text{ and} \\ L_\infty(\lambda) &= \sup \left|\hat{\lambda}-\lambda\right|. \end{aligned}$$

For an approach using the L_1 metric see [10]. In this thesis we will use the L_2 error criterion for its technical tractability . The Integrated Square Error (ISE) is defined by

$$\text{ISE}\left(\hat{\lambda}(x)\right) = \int \left(\hat{\lambda}(x)-\lambda(x)\right)^2 w(x)\,dx,$$

where $w(x)$ is an appropriately defined weight function. Note that the above error criterion leads to a random quantity. A non-random error criterion can be defined by considering the expected value of any random criterion of accuracy. In particular, taking the expected value of the squared error leads to the definition of the Mean Squared Error (MSE), i.e.

$$\text{MSE}\left(\hat{\lambda}(x)\right) = \mathbb{E}\left(\lambda(x)-\hat{\lambda}(x)\right)^2$$

which assess the performance of $\lambda(x)$ at the point x. Expanding the square on the RHS of the above equation,

$$\text{MSE}\left(\hat{\lambda}(x)\right) = \left\{\mathbb{E}\hat{\lambda}(x)-\lambda(x)\right\}^2 + \mathbb{V}\text{ar}\left(\hat{\lambda}(x)\right).$$

A global non random measure of accuracy can then be defined by integrating the MSE. This leads to the definition of the Mean Integrated Squared Error (MISE), i.e.

$$\text{MISE}\left(\hat{\lambda}(x)\right) = \mathbb{E}\int \left(\hat{\lambda}(x)-\lambda(x)\right)^2 w(x)\,dx$$

where $w(x)$ is a nonnegative weight function. The MSE properties of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ have been studied by many authors. We summarize their results in the following theorem. Firstly, note that $\sup_x |F_n(x)-F(x)| = O_p(1/\sqrt{n})$. This means that asymptotically $\hat{\lambda}_1(x|h)$ and $\hat{\lambda}_2(x|h)$ are equivalent to

$$\begin{aligned} \bar{\lambda}_1(x|h) &= \frac{\hat{f}(x|h)}{1-F(x)} \\ \text{and} \quad \bar{\lambda}_2(x|h) &= \frac{1}{n}\sum_{i=1}^{n} \frac{K_h(x-X_i)}{1-F(X_i)}. \end{aligned}$$

Then we have the following theorem

Theorem 1.3.1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with common distribution F and density f, which is twice continuously differentiable. Let $n \to +\infty$, $h \to 0$ and $nh \to \infty$. Also, suppose that there exists small enough h such that $K_h(y-x)/(1-F(y))$ is uniformly bounded for $|y-x| > M$, for any $M > 0$. Then*

$$\begin{aligned}\mathbb{E}\bar{\lambda}_1(x|h) - \lambda(x) &= \frac{f''(x)}{1-F(x)}\frac{h^2}{2}\int u^2 K(u)\,du + o(h^2),\\ \mathbb{V}\mathrm{ar}\left(\bar{\lambda}_1(x|h)\right) &= \frac{1}{nh}\frac{\lambda(x)}{1-F(x)}\int K^2(u)\,du + o\left(\frac{1}{nh}\right)\end{aligned}$$

and thus

$$MSE\left(\bar{\lambda}_1(x|h)\right) = h^4\mu_2^2\left(\frac{f''(x)}{1-F(x)}\right)^2 + \frac{1}{nh}\frac{\lambda(x)}{1-F(x)}\int K^2(u)\,du + o\left(h^4\right) + o\left(\frac{1}{nh}\right)$$

with $\mu_2 = \int (u^2/2)K(u)\,du$. Also

$$\begin{aligned}\mathbb{E}\bar{\lambda}_2(x|h) - \lambda(x) &= \lambda''(x)\frac{h^2}{2}\int u^2 K(u)\,du + o(h^2),\\ \mathbb{V}\mathrm{ar}\left(\bar{\lambda}_2(x|h)\right) &= \frac{1}{nh}\frac{\lambda(x)}{1-F(x)}\int K^2(u)\,du + o\left(\frac{1}{nh}\right)\end{aligned}$$

hence,

$$MSE\left(\bar{\lambda}_2(x|h)\right) = h^4\mu_2^2\lambda''(x)^2 + \frac{1}{nh}\frac{\lambda(x)}{1-F(x)}\int K^2(u)\,du + o\left(h^4\right) + o\left(\frac{1}{nh}\right).$$

Proof. For a proof see either [34] or [49]. ∎

Remark 1.1. The presence of the survival function, $1-F(x)$, in the denominator of the variance component of the MSE for both estimators means that for x large the variance of both estimators will increase.

Remark 1.2. As it has been noted in [48], calculation of the exact MSE of $\hat{\lambda}_1$ is intractable. However it is possible to find the mean and variance of $\hat{\lambda}_2$. Notice that,

$$\begin{aligned}\mathbb{E}\left(\frac{K_h\left(x-X_i\right)}{1-F_n(X_i)}\right) &= \mathbb{E}\left\{\mathbb{E}\left(\frac{K_h\left(x-X_i\right)}{1-F_n(X_i)}\middle| X_i = r\text{th order statistic}\right)\right\}\\ &= n\mathbb{E}\left\{\mathbb{E}\left(\frac{K_h\left(x-X_{(r)}\right)}{n-r+1}\right)\right\}\\ &= \mathbb{E}\int n\frac{K_h\left(x-u\right)}{n-r+1}\frac{n!}{(r-1)!(n-r)!}F^{r-1}(u)(1-F(u))^{n-r}f(u)\,du.\end{aligned}$$

Since $r = j$ with probability $1/n$,

$$\begin{aligned}\mathbb{E}\left(\frac{K_h\left(x-X_i\right)}{1-F_n(X_i)}\right) &= \frac{1}{n}\sum_{j=1}^{n}\int nK_h\left(x-u\right)\binom{n}{j-1}F^{j-1}(u)(1-F(u))^{n-j}f(u)\,du\\ &= \int K_h\left(x-u\right)\left\{\frac{1-F^n(u)}{1-F(u)}\right\}f(u)\,du.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left\{\hat{\lambda}_2(x|h)\right\} &= \sum_{i=1}^{n} \frac{1}{n} \int K_h(x-u) \left\{\frac{1-F^n(u)}{1-F(u)}\right\} f(u)\,du \\ &= \int K_h(x-u)\,\lambda(u)\,du - \int K_h(x-u)\,\lambda(u) F^n(u)\,du.\end{aligned}$$

Performing the change of variables $x - u = ht$, expanding $\lambda(x - ht)$ in Taylor series around x and noting that asymptotically $F^n(u) = o(n^{-1})$ together with the standard assumptions on the kernel that it integrates to 1 and that the integral of $zK'(z)$ is 0 gives the bias formula for $\hat{\lambda}_2$ of theorem 1.3.1. The variance is

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left\{\hat{\lambda}_2(x|h)\right\} = \int I_n(F_n(u))\lambda(u)K_h^2(x-u)\,du& \\ + 2\iint_{u\le z} A_n(u,z)K_h(x-u)K_h(x-z)\lambda(u)\lambda(z)\,du\,dz&,\end{aligned}$$

where

$$\begin{aligned}I_n(F(u)) &= \sum_{i=1}^{n} \frac{1}{n-i+1}\binom{n}{i-1} F(u)^{i-1}(1-F(u))^{n-i+1}, \\ A_n(u,z) &= (1-F^n(u))F^n(z) - (1-F(u))\frac{F^n(z)-F^n(u)}{F(z)-F(u)}.\end{aligned}$$

It has been proven in [49] that asymptotically, $nI_n(F(u))$ converges to $1/(1-F(u))$ and that the term A_n is negligible. This, together with the change of variables $x - u = ht$, a Taylor expansion of $\lambda(x-ht)/(1-F(x-ht))$ around x and the same assumptions on the kernel that were used for the bias, gives the variance expression of $\hat{\lambda}_2$ in theorem 1.3.1.

The literature on asymptotic results for both estimators is quite rich. Rice and Rosenblatt in [34] prove that asymptotically $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are equivalent to $\bar{\lambda}_1$. In the same paper they obtain the result on the maximal weighted deviation between an estimate of the failure rate and the true hazard and also prove the asymptotic normality of $\bar{\lambda}_1$. Tanner and Wong in [44] establish the asymptotic normality of estimator $\hat{\lambda}_2$ by using the Hajek projection method. The same result was achieved by Lo Mack and Wang in [22] via the strong representation of the Kaplan-Mayer estimator. Patil *et al.* in [30], assuming a martingale point of view showed that asymptotically $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are equivalent with $\hat{\lambda}_2$ being more natural as a choice.

1.4 Bandwidth selection.

The critical role that the bandwidth plays in the performance of the estimators is quantified mathematically in theorem 1.3.1. First, note that in both MSE expressions above the first terms are squared bias contributions while the second terms are variance contributions. Asymptotically the squared bias is proportional to h^4, so for this quantity to decrease one has to take h small. However, as can be seen in figure 1.1, this will lead to an increase of the leading term of the variance since this quantity is asymptotically proportional to $(nh)^{-1}$. Thus, in order to keep the MSE as small as possible, h should be chosen in such a way so that the two terms of the MSE are minimal. The importance of bandwidth to the performance of kernel based estimators has led to the development of many methods as it can be seen in the literature. Here we give an introduction to the methods which will be used in the present thesis.

1.4.1 Direct plug-in rules.

Due to the similarities in both estimators introduced so far we will demonstrate some of the most common bandwidth selection methods using a unified approach. Let $q(x)$ be either $f''(x)/(1-F(x))$ or $\lambda''(x)$ and $\bar{\lambda}(x|h)$ be either $\bar{\lambda}_1(x|h)$ or $\bar{\lambda}_2(x|h)$. For simplicity assume that λ'' and f'' are square integrable. Then the MSE of $\bar{\lambda}(x|h)$ is

$$\text{MSE}\left(\bar{\lambda}(x|h)\right) = h^4\mu_2^2(K)q^2(x) + \frac{1}{nh}\frac{\lambda(x)}{1-F(x)}\int K^2(z)\,dz + o\left(\frac{1}{nh}\right) + o\left(h^4\right).$$

If the aim is to find the best bandwidth for the whole curve, then h should be selected to minimize the MISE. Integrating, and denoting with $R(g) = \int g^2(x)\,dx$ gives

$$\text{MISE}\left(\bar{\lambda}(x|h)\right) = h^4\mu_2^2(K)\int q^2(x)w(x)\,dx + \frac{1}{nh}R(K)\int \frac{\lambda(z)}{1-F(z)}w(z)\,dz + o\left(\frac{1}{nh}\right) + o\left(h^4\right).$$

Now, rewrite this expression as

$$\text{MISE} = \text{AMISE} + o\left(\frac{1}{nh}\right) + o\left(h^4\right)$$

where

$$\text{AMISE}\left(\bar{\lambda}(x|h)\right) = h^4\mu_2^2(K)\int q^2(x)w(x)\,dx + \frac{1}{nh}R(K)\int \frac{\lambda(z)}{1-F(z)}w(z)\,dz.$$

We call this quantity asymptotic MISE (AMISE). Since it provides a good large sample approximation of the MISE and since it is a much easier expression to comprehend compared to that of the MISE, we consider minimization of the AMISE. It is easy to see that the minimizer of AMISE is

$$h_{\text{AMISE}} = \left\{\frac{R(K)M}{n\mu_2^2(K)S(q(x))}\right\}^{\frac{1}{5}}$$

with M and $S(q)$ being

$$M = \int \frac{\lambda(z)}{1-F(z)}w(z)\,dz \quad \text{and} \quad S(q) = \int q^2(x)w(x)\,dx.$$

Asymptotically, if the approximation of the MISE by the AMISE is good enough, then h_{AMISE} is expected to behave well for the MISE as well. By substituting the unknown quantities by their estimators in the above formulas we get the so-called 'direct plug-in' rule. A problem here is that we will need to use an initial (pilot) bandwidth for the kernel estimate of the functional q. This can be remedied by using a quick and simple method such as that described in section 3.2.1 of [47].

1.4.2 Least squares cross validation.

Among the earliest fully automatic bandwidth selection techniques is the least squares cross validation method. It is based on minimization of the ISE of an estimator of the hazard rate. In the case of $\bar{\lambda}$, from the definition of the ISE we have,

$$\mathrm{ISE}(\bar{\lambda}, h) = \int \bar{\lambda}^2(x|h)w(x)\,dx - 2\int \bar{\lambda}(x|h)\lambda(x)w(x)\,dx + \int \lambda^2(x)w(x)\,dx.$$

Since the third term on the RHS of the above equation does not depend on h it is equivalent to consider only minimization of the first two. The problem is that the second term includes the unknown hazard rate. However, since

$$\int \bar{\lambda}(x|h)\lambda(x)w(x)\,dx = \int \frac{\bar{\lambda}(x|h)}{1-F(x)}f(x)w(x)\,dx$$

the second term can be replaced by a nearly unbiased estimator, such as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\bar{\lambda}^{-i}(X_i|h)}{1-F_n(X_i)}w(X_i)$$

where $\bar{\lambda}^{-i}$ denotes the leave-one-out version of $\bar{\lambda}$. In the case of $\bar{\lambda}_1$ it is given by

$$\bar{\lambda}_1^{-i}(x) = \frac{\hat{f}^{-i}(x|h)}{1-F(x)}$$

where $\hat{f}^{-i}$ is the kernel estimate of f based on the sample $X_1, \ldots X_{i-1}, X_{i+1}, \ldots X_n$. In the case of $\bar{\lambda}_2$,

$$\bar{\lambda}_2^{-i}(x|h) = \frac{1}{n-1}\sum_{j\neq i}^{n}\frac{K_h\left(x-X_j\right)}{1-F(X_j)}.$$

This defines the least squares cross validation criterion

$$CV(h) = \int \bar{\lambda}^2(x|h)\,dx - \frac{2}{n}\sum_{i=1}^{n}\frac{\bar{\lambda}^{-i}(X_i|h)}{1-F_n(X_i)}w(X_i)$$

which is minimized to find the data-based bandwidth choice. For detailed work on least squares cross validation in nonparametric hazard rate estimation settings we refer the reader to Sarda and Vieu [36] and Patil [28], [29].

1.5 Problems with kernel-based hazard rate estimators.

Although kernel-based estimators have an intuitive appeal, there are potential difficulties caused by individual characteristics of the curve under estimation. In this section we describe two important problems that are addressed in this dissertation. In subsection 1.5.1 we explain the problem of boundary bias and in subsection 1.5.2 we give a precise description of the usual bias problem in nonparametric hazard rate estimation.

1.5.1 Boundary bias.

A special form of bias, called *boundary bias*, occurs when an estimator is biased more than usual towards the boundary (or the boundaries) of the region of estimation. To give more insight into this situation consider for instance $\bar{\lambda}_1$. Note that the support of $\bar{\lambda}_1$ practically is $[0, \max\{X_i\} + h]$ because at the largest observation the kernel will take into account observations in the range $(\max X_i - h, \max X_i + h)$. If we take the support of the function λ to be the interval $[0,1]$ then the expected value of $\bar{\lambda}_1$ at 0 is approximately half the actual value of λ as will be seen shortly.

Suppose that we estimate λ at some x within the boundary region. That is $x = ph$, $0 \le p < 1$. Consider any kernel defined on $(-1,1)$ and let

$$a_l(p) = \int_{-1}^{p} u^l K(u)\, du \;\text{ and }\; b(p) = \int_{-1}^{p} K^2(u)\, du.$$

Then

$$\mathbb{E}\bar{\lambda}_1(x|h) = \frac{1}{1-F(x)} \int_0^1 K_h(x-y) f(y)\, dy = \frac{1}{h} \int_0^{h(1+p)} K\left(p - \frac{y}{h}\right) \frac{f(y)}{1-F(x)}\, dy.$$

Set $x - y = hu$. Then, for $x + h < y < 0$, $u \in (-1, p)$ and $du = -h\, dy$. Thus, expanding $f(x)$ in Taylor series around x in the second step below,

$$\begin{aligned}
\mathbb{E}\bar{\lambda}_1(x|h) &= \int_{-1}^{p} \frac{K(u)}{1-F(x)} f(x-hu)\, du \\
&= \int_{-1}^{p} \frac{K(u)}{1-F(x)} \left\{ f(x) - huf'(x) + \frac{(hu)^2}{2!} f''(x) - \dots \right\} du \\
&= a_0(p)\lambda(x) - a_1(p)h\frac{f'(x)}{1-F(x)} + \frac{h^2}{2!} a_2(p) \frac{f''(x)}{1-F(x)} + o(h^2).
\end{aligned}$$

At $x = 0$, $a_0(0) = 1/2$, so $\mathbb{E}\bar{\lambda}_1(0) \simeq \frac{1}{2}\lambda(0)$. Similarly the boundary bias for $\bar{\lambda}_2$ will be

$$\mathbb{E}\hat{\lambda}_2(x) = a_0(p)\lambda(x) - a_1(p)h\lambda'(x) + \frac{h^2}{2!} a_2(p)\lambda''(x) + o(h^2).$$

That is, in both cases the bias at the boundary is of order $O(h)$ or more in contrast to the $O(h^2)$ rate inside the interior. Of course if $p = 1$ then $a_0(1) = 1$, $a_1(1) = 0$ and $a_2(1) = \int u^2 K(u)\, du$. Thus the bias expressions above are the usual interior bias expressions of theorem 1.3.1.

A possible solution for this problem is to use special forms of kernels, known as boundary kernels, which are weight functions that are used only within the boundary region. Let $c_l(p) = \int_{-1}^{p} u^l L(u)\, du$ with L being a kernel function different from, but related to, K. Let $\bar{\lambda}_{2,K}$ be the $\bar{\lambda}_2$ estimate constructed by using K and $\bar{\lambda}_{2,L}$ the same estimate based on L. Since

$$\mathbb{E}\left(c_1(p)\bar{\lambda}_{2,K}(x)\right) = c_1(p)a_0(p)\lambda(x) - ha_1(p)c_1(p)\lambda'(x) + O(h^2)$$

and

$$\mathbb{E}\left(a_1(p)\bar{\lambda}_{2,L}(x)\right) = a_1(p)c_0(p)\lambda(x) - ha_1(p)c_1(p)\lambda'(x) + O(h^2)$$

a linear combination of the two kernels

$$B(x) = \frac{c_1(p)K(x) - a_1(p)L(x)}{c_1(p)a_0(p) - a_1(p)c_0(p)}$$

yields

$$\mathbb{E}\left(\tilde{\lambda}_B(x)\right) = \lambda(x) + O(h^2).$$

where $\bar{\lambda}_B$ is estimator $\bar{\lambda}_2$ based on the kernel B. Thus, boundary and interior bias are of the same order. In the present thesis we develop an approach based on the method of local linear fitting. This method has the advantage that it adapts automatically to boundary points and as a result it produces estimators that have the same amount of boundary bias as in areas away from the boundaries.

1.5.2 Bias reduction.

From theorem 1.3.1 we have that the squared bias of $\hat{\lambda}_2$ is $O(h^4)$ and the variance is $O((nh)^{-1})$. Clearly a bandwidth of size $O(n^{-\frac{1}{5}})$ leads to MSE of order $O(n^{-\frac{4}{5}})$. Under the current settings Singpurwalla and Wong in [41] proved that the rate of convergence of the MSE is at most $O(n^{-\frac{4}{5}})$. To improve the accuracy, that is, to increase the rate of convergence of the MSE one way is to consider a method that reduces the bias. Among the earliest methods employed for this purpose is the 'higher order kernels' method which we briefly describe below. A higher order kernel of order k is a kernel such that

$$\int u^i K(u)\, du = \begin{cases} 0 & i = 1, \dots, k-1 \\ a \neq 0 & i = k \\ b_i \geq 0 & i = k+1, \dots \end{cases}$$

From the analysis of the bias for estimator $\hat{\lambda}_2$ in remark 1.2, and provided that λ is sufficiently smooth and that a symmetric kernel K is used, it follows that the asymptotic bias of $\hat{\lambda}_2$ has a formal expansion of the form

$$\mathbb{E}\hat{\lambda}_2(x) - \lambda(x) = \sum_{i=1}^{+\infty} \frac{h^{2i}}{(2i)!}\lambda^{(2i)}(x) \int u^{2i} K(u)\, du + o(n^{-1}). \tag{1.4}$$

Then, it can be seen immediately that by the use of a higher order kernel the bias can be reduced asymptotically to any desired order. However, this comes at a price. The use of higher order kernels creates estimators that produce negative values at the tails, and that is something undesirable for estimates of a nonnegative valued function such as the hazard rate function.

A closer look at the estimation procedure reveals that the possible reason for higher bias is that the use of constant bandwidth results in the estimator not adapting to local variations in the curve. It is therefore desirable to generalize the estimation procedure so that different degrees of smoothing are applied according to whether the region of estimation is of relatively low or high density. This can be done by allowing the bandwidth to vary either according to data points, leading to variable bandwidth hazard estimators or according to the location of estimation, leading to variable kernel hazard estimators.

Using as reference estimator $\hat{\lambda}_2$ a general formula for variable kernel estimators is

$$\hat{\lambda}(x) = \frac{1}{h(x)} \sum_{i=1}^{n} \frac{K\left(\frac{x-X_{(i)}}{h(x)}\right)}{n-i+1}.$$

Obviously, choice of $h(x)$ is crucial for the performance of the method. The most common choice for $h(x)$ is the kth nearest neighbor distance from the data points to x. Essentially the method leads to $h \propto 1/\lambda(x)$. This approach was taken by McCune and McCune in [25] yielding a nonnegative estimator with asymptotic MSE equal to $O(n^{-\frac{8}{9}})$. However it has to be noted ([40], pp. 58) that there are situations where $h \propto 1/\lambda(x)$ is unsatisfactory and as a consequence the resultant estimate does not provide the desired improved local adaptivity.

A general formulation of variable bandwidth estimators, in the case of $\hat{\lambda}_2$ is

$$\hat{\lambda}(x) = \sum_{i=1}^{n} \frac{1}{h(X_{(i)})} \frac{K\left(\frac{x-X_{(i)}}{h(X_{(i)})}\right)}{n-i+1}.$$

Motivated by the theoretical work in [2] for the density case, in this thesis we extend $\hat{\lambda}_1$ and $\hat{\lambda}_2$ so that they use variable bandwidth. We then show that variable bandwidth yields the same advantages as it does in the density case.

In both cases it has to be noted that practical implementation of the methods requires specification of the bandwidth law which depends on the real curve. This reduces its practical usefulness. However, in this thesis we first study these *ideal* versions of the variable bandwidth estimators and we show that they have bias of order h^4. We then develop practically feasible estimators, by replacing the unknowns in the bandwidth law formulae by suitable fixed bandwidth kernel estimates.

1.5.3 Transformations.

Notice from (1.4) that the bias will be completely eliminated if $\lambda^{(2i)} = 0$ for every i. This will be the case if $\lambda(x)$ is a constant function. Therefore, if X is exponentially distributed then one expects to have zero bias for $\hat{\lambda}_2$ except in the boundaries. We make use of this idea to propose a transformation based method to reduce the bias. This proposal is analogous to the transformation based method of bias reduction in density settings, see for example [35]. From the theorem of transformations of random variables we have

$$\lambda_X(x) = \lambda_Y(g(x))g'(x)$$

where λ_X and λ_Y denote the hazard rate of X and Y respectively and $g(X) = Y$ is the one to one function used to transform the sample. Thus, we first transform the random sample X_i to Y_i, $i = 1, \ldots, n$ such that $Y_i = g(X_i)$ is exponentially distributed. Then, using the definition of $\hat{\lambda}_2$ we obtain the estimate $\hat{\lambda}_Y$ of λ_Y, and finally we transform back $\hat{\lambda}_Y$ to obtain the estimate of λ_X as

$$\hat{\lambda}_X(x) = \hat{\lambda}_Y(g(x))g'(x) = \frac{g'(x)}{nh} \sum_{i=1}^{n} \frac{K\left(\frac{g(x)-g(X_i)}{h}\right)}{1-F_n(g(X_i))}$$

Note here that although this estimate belongs to neither class of estimates defined in the previous subsection, applying the mean value theorem and assuming that g is a monotonically increasing function we get,

$$\hat{\lambda}_X(x) = \frac{g'(x)}{h} \sum_{i=1}^{n} \frac{K\left(\frac{x-X_i}{h} g'(\xi_i)\right)}{n-i+1}, \quad \xi_i \in [x, X_i],$$

which reveals that this estimate has something in common with the above estimates. Quite naturally the most important issue here is the choice of the transformation. It mostly depends on the shape of the hazard rate, and both parametric and nonparametric approaches are possible. In this thesis we develop a purely nonparametric method for estimating the hazard rate via transformations based on the $\hat{\lambda}_2$ estimator.

1.6 Outline of the thesis.

The problem of the boundary bias is examined in the second chapter. There, we develop local linear fit based estimators for the hazard rate. They have the property that they adapt readily to boundary points. This means that the resulting estimators in the boundary have the same amount of bias as in the interior.

In the third chapter we begin by extending estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ to the point that they make use of the advantages offered by applying different amounts of smoothing at different places. We then study their asymptotic behavior and prove that in this situation there is an improvement in the mean integrated square error of order to $n^{-\frac{8}{9}}$ over the standard second order kernel.

In the fourth chapter we extend the method of transformations to the hazard rate case. The asymptotic study of the squared error of the resulting estimator shows that the improvement over the traditional kernel hazard estimators is similar to the variable bandwidth method. We then proceed with some graphical illustrations of the last two methods.

Finally, in the fifth chapter we present a summary of the results and we suggest topics for future research.

Chapter 2

HAZARD RATE ESTIMATION USING LOCAL LINEAR FIT

2.1 Introduction.

We have stated the problem of boundary bias of kernel hazard rate estimators precisely in section 1.5.1. In this chapter we address this issue. The usual remedy for boundary bias is to use specially developed boundary kernels (see for example [40]). This approach requires complicated adjustments to the estimator and so it is desirable to have an estimator that handles boundary effects correctly in the sense that it achieves the same asymptotic bias rate of convergence as in the interior.

The local linear fit method was employed by Fan and Gijbels in [13] to address the problem of boundary bias in the density estimation setting. There, it is observed that the method readily adapts to boundary points and produces estimators with the same amount of boundary bias as in places outside the boundary. Enthused by the performance of the method we propose to bring in the technique to the hazard rate case.

Suppose that we are interested in estimating the hazard function $\lambda(x)$ on an interval $[0, b]$ of the real line, with $b > 0$, given a sample $X_1, \ldots, X_N$ from some density f. The first step is to discretize the data. Partition the interval into n subintervals $\{I_i, i = 1 \ldots n\}$ of equal length $\Delta = b/n$ and let x_i, $i = 1 \ldots n$, be the center of the interval I_i. If we can construct appropriate data-based estimators for the hazard rate, say Y_i, at each bin center x_i such that

$$\mathbb{E}Y_i \simeq \lambda(x_i), \quad \mathbb{V}\mathrm{ar}(Y_i) \simeq \sigma(x_i),$$

then the problem can be formulated as a fixed design nonparametric regression problem based on the approximately independent data (x_i, Y_i),

$$Y_i = m(x_i) + \sigma(x_i)\varepsilon_i, \quad i = 1, 2, \ldots, n \tag{2.1}$$

with $m(x_i) = \lambda(x_i)$ and ε_i being independent r.v. with mean 0 and variance 1.

The literature (see for example Elandt-Johnson and Johnson [11] and Cox and Oakes [9]) gives us a number of ways to construct empirical estimators of the hazard rate. The most natural one is the relative frequency estimator which is constructed by dividing the relative frequency by the empirical survival function. This results in a 'histogram' type of estimator, Y_i, with the value of the estimator being constant in each bin i. By choosing the bin center, x_i, as the point of estimation we could treat the hazard estimation problem as an heteroscedastic nonparametric regression problem based on the data $\{x_i, Y_i, i = 1, \ldots, n\}$.

In section 2.3 we turn our attention to estimation of the regression function $m(x)$. The method we use here is local polynomial fitting which approximates the unknown curve by fitting a polynomial of order p in a neighborhood of x to the data set $\{x_i, Y_i, i = 1, \ldots, n\}$ using weighted least squares. Fan and Gijbels in [13] recommend using p=$\nu + 1$ as the order of the polynomial if interest is to estimate the νth derivative of the hazard rate. Thus, for estimation of the hazard function λ, where $\nu = 0$, we take $p = 1$. This leads to local linear fitting, that is, we approximate λ by fitting straight lines in local neighborhoods. The asymptotic properties of the derived estimator are studied in section 2.4.

The bandwidth used for the local least squares fits controls the neighborhood of smoothing. Therefore, bandwidth choice plays a crucial role in practical implementation of the estimators. In section 2.5, with practical implementation being the primary objective, we discuss two approaches to bandwidth selection, a simple plug-in method based on minimization of the MISE of the hazard rate estimators and a cross validation method which gives ISE optimal bandwidth. Several interesting theoretical issues are associated with both approaches and will be addressed in future work.

2.2 An empirical estimator of the hazard rate.

Associated with each x_i, the center of the interval $I_i, i = 1, 2, \ldots, n$, define

$$Y_i = \frac{\frac{\{\#\text{ of } X_i \in I_i\}}{N\Delta}}{1 - \sum_{j=1}^{i} \frac{\{\#\text{ of } X_i \in I_j\}}{N}} = \frac{f_i/\Delta}{N - \sum_{j=1}^{i} f_j} \quad \text{with} \quad f_i = \{\#\text{ of } X_i \in I_i\}.$$

Note that if $i = n$ then

$$N - \sum_{j=1}^{n} f_j = 0.$$

In order to avoid division by zero, we redefine the Y_i's to be

$$Y_i = \frac{1}{\Delta} \frac{f_i}{N - \sum_{j=1}^{i} f_j + 1}, \quad i = 1, 2, \ldots, n.$$

Also, let

$$c_i = \frac{f_i}{N - \sum_{j=1}^{i} f_j + 1}, \quad i = 1, 2, \ldots, n.$$

Note that,

$$f_i \sim \text{binomial}\,(N, p_i)\,, \;\; p_i = \int_{I_i} f(x)\,dx, \;\text{ and }\; F_i \equiv \sum_{j=1}^{i} f_j \sim \text{binomial}\,(N, P_i)$$

where

$$P_i = \sum_{j=1}^{i} p_j.$$

Now,

$$\mathbb{E}\{c_i\} = \mathbb{E}\left(\mathbb{E}\left\{\frac{f_i}{N - \sum_{j=1}^{i} f_j + 1}\middle| \sum_{j=1}^{i} f_j = m\right\}\right) = \mathbb{E}\left(\mathbb{E}\left\{\frac{f_i}{N-m+1}\middle| \sum_{j=1}^{i} f_j = m\right\}\right).$$

Note that conditional on $\sum_{j=1}^{i} f_j = m$ the distribution of f_i is binomial$(m, p_i/P_i)$. Since $\mathbb{E}f_i = mp_i/P_i$, the mean of c_i becomes

$$\mathbb{E}\{c_i\} = \frac{p_i}{P_i}\mathbb{E}\left\{\frac{m}{N-m+1}\right\}$$

where now m itself is a binomial random variable with parameters N and P_i. Thus,

$$\begin{aligned}
\mathbb{E}\left\{\frac{m}{N-m+1}\right\} &= \sum_{m=0}^{N} \frac{m}{N-m+1}\binom{N}{m} P_i^m (1-P_i)^{N-m} \\
&= \frac{P_i}{1-P_i}\sum_{m=1}^{N}\binom{N}{m-1} P_i^{m-1}(1-P_i)^{N-m+1} \\
&= \frac{P_i}{1-P_i}\left\{\sum_{r=0}^{N}\binom{N}{r} P_i^r(1-P_i)^{N-r} - P_i^N\right\}, \quad \text{for } r = m-1 \\
&= \frac{P_i}{1-P_i}\left(1-P_i^N\right) = \frac{P_i}{1-P_i} + o\left(\frac{1}{N}\right). \qquad (2.2)
\end{aligned}$$

That is,

$$\mathbb{E}\{c_i\} = \frac{p_i}{P_i}\left\{\frac{P_i}{1-P_i} + o\left(\frac{1}{N}\right)\right\} = \frac{p_i}{1-P_i} + o\left(\frac{1}{N}\right). \qquad (2.3)$$

Now, consider the variance of c_i. We have,

$$\mathbb{V}\mathrm{ar}\{c_i\} = \mathbb{E}\left\{\mathbb{V}\mathrm{ar}\left(\frac{f_i}{N-\sum_{j=1}^{i} f_j+1}\middle| \sum_{j=1}^{i} f_j = m\right)\right\} + \mathbb{V}\mathrm{ar}\left\{\mathbb{E}\left(\frac{f_i}{N-\sum_{j=1}^{i} f_j+1}\middle| \sum_{j=1}^{i} f_j = m\right)\right\}.$$

Since for given $\sum_{j=1}^{i} f_j = m$, we have

$$\mathbb{V}\mathrm{ar}\{f_i\} = m\frac{p_i}{P_i}\left(1-\frac{p_i}{P_i}\right) \quad \text{and} \quad \mathbb{E}f_i = \frac{mp_i}{P_i}$$

then,

$$\mathbb{V}\mathrm{ar}\{c_i\} = \frac{p_i(P_i - p_i)}{P_i^2}\mathbb{E}\left\{\frac{m}{(N-m+1)^2}\right\} + \frac{p_i^2}{P_i^2}\mathbb{V}\mathrm{ar}\left\{\frac{m}{N-m+1}\right\}. \tag{2.4}$$

Further,

$$\begin{aligned}\mathbb{E}\left\{\frac{m}{(N-m+1)^2}\right\} &= \sum_{m=1}^{N}\frac{1}{N-m+1}\binom{N}{m-1}P_i^m(1-P_i)^{N-m}\\ &= \frac{P_i}{1-P_i}\sum_{m=1}^{N}\frac{1}{N-m+1}\binom{N}{m-1}P_i^{m-1}(1-P_i)^{N-m+1}\\ &= \frac{P_i}{1-P_i}I_N(P_i) = \frac{1}{N}\frac{P_i}{(1-P_i)^2} + o\left(\frac{1}{N}\right),\end{aligned} \tag{2.5}$$

since, as $N \to +\infty$,

$$NI_N(P_i) \to \frac{1}{1-P_i}.$$

Also note that

$$\begin{aligned}\mathbb{E}\left\{\frac{m(m-1)}{(N-m+1)^2}\right\} &= N\sum_{m=2}^{N}\frac{(N-1)!}{(m-2)!(N-m+1)!}\frac{1}{N-m+1}P_i^m(1-P_i)^{N-m}\\ &= N\frac{P_i^2}{1-P_i}\sum_{m=2}^{N}\binom{N-1}{m-2}\frac{1}{N-m+1}P_i^{m-2}(1-P_i)^{N-m+1}\\ &= N\frac{P_i^2}{1-P_i}\sum_{r=1}^{N}\binom{N-1}{r-1}\frac{1}{N-r}P_i^{r-1}(1-P_i)^{N-r}, \quad \text{for } r=m-1\\ &= N\frac{P_i^2}{1-P_i}I_{N-1}(P_i) = \frac{N}{N-1}\frac{P_i^2}{1-P_i}(N-1)I_{N-1}(P_i)\\ &= \frac{P_i^2}{(1-P_i)^2} + o\left(\frac{1}{N}\right).\end{aligned} \tag{2.6}$$

Therefore, using (2.2), (2.5) and (2.6) we get

$$\begin{aligned}\mathbb{V}\mathrm{ar}\{c_i\} &= \frac{p_i(P_i-p_i)}{P_i^2}\frac{1}{N}\frac{P_i}{(1-P_i)^2}\\ &\quad + \frac{p_i^2}{P_i^2}\left\{\frac{P_i^2}{(1-P_i)^2} + \frac{1}{N}\frac{P_i}{(1-P_i)^2} - \frac{P_i^2}{(1-P_i)^2}\right\} + o\left(\frac{1}{N}\right)\\ &= \frac{1}{N}\frac{p_i(P_i-p_i)}{P_i(1-P_i)^2} + \frac{1}{N}\frac{p_i^2}{P_i(1-P_i)^2} + o\left(\frac{1}{N}\right)\\ &= \frac{1}{N}\frac{p_i}{(1-P_i)^2} + o\left(\frac{1}{N}\right).\end{aligned} \tag{2.7}$$

Thus, for sufficiently small Δ from (2.3) and (2.7) we have

$$\mathbb{E}\{Y_i\} = \mathbb{E}\left\{\frac{1}{\Delta}c_i\right\} \simeq \lambda(x_i) \;\text{ and }\; \mathbb{V}\mathrm{ar}\{Y_i\} = \mathbb{V}\mathrm{ar}\left\{\frac{1}{\Delta}c_i\right\} \simeq \frac{1}{N\Delta}\frac{\lambda(x_i)}{1-F(x_i)}. \tag{2.8}$$

Thus the data $\{x_i, Y_i, i = 1, \ldots, n\}$ can be used to provide a hazard rate estimator based on the local linear method. Next we calculate $\mathbb{E}(c_i c_j)$ which will be used in section 2.4. We have,

$$\mathbb{E}\{c_i c_j\} = \mathbb{E}\left\{\frac{f_i}{N - \sum_{k=1}^{i} f_k + 1} \frac{f_j}{N - \sum_{k=1}^{j} f_k + 1}\right\}.$$

Without loss of generality assume $i < j$. For given $\sum_{k=1}^{i} f_k = m_1$ and $\sum_{k=i+1}^{j} f_k = m_2$ we have that

$$f_i \sim \text{binomial}\left(m_1, \frac{p_i}{P_i}\right) \quad \text{and} \quad f_j \sim \text{binomial}\left(m_2, \frac{p_j}{P_j - P_i}\right).$$

Therefore,

$$\begin{aligned}\mathbb{E}\{c_i c_j\} &= \mathbb{E}\left\{\mathbb{E}\left(\frac{f_i}{N - m_1 + 1} \frac{f_j}{N - m_1 - m_2 + 1}\right)\right\} \\ &= \frac{p_i p_j}{P_i(P_j - P_i)} \mathbb{E}\left\{\frac{m_1 m_2}{(N - m_1 + 1)(N - m_1 - m_2 + 1)}\right\}.\end{aligned}$$

Now, given m_1, $m_2 \sim \text{binomial}(N - m_1, (P_j - P_i)(1 - P_i)^{-1})$ and so

$$\mathbb{E}\left\{\frac{m_1 m_2}{(N - m_1 + 1)(N - m_1 - m_2 + 1)}\right\} = \mathbb{E}\left\{\frac{m_1}{N - m_1 + 1} \mathbb{E}\left(\frac{m_2}{N - m_1 - m_2 + 1}\right)\right\}.$$

Now, let $N - m_1 = N_1$ and $P = (P_j - P_i)(1 - P_i)^{-1}$. Then,

$$\begin{aligned}\mathbb{E}\left(\frac{m_2}{N_1 - m_2 + 1}\right) &= \sum_{m_2=0}^{N_1} \frac{m_2}{N_1 - m_2 + 1} \binom{N_1}{m_2} P^{m_2} (1 - P)^{N_1 - m_2} \\ &= \frac{P_j - P_i}{1 - P_j}\left\{1 - \left(\frac{P_j - P_i}{1 - P_i}\right)^{N - m_1}\right\}.\end{aligned}$$

Thus,

$$\mathbb{E}\left\{\frac{m_1 m_2}{(N - m_1 + 1)(N - m_1 - m_2 + 1)}\right\} =$$

$$= \frac{P_j - P_i}{1 - P_j} \mathbb{E}\left\{\frac{m_1}{N - m_1 + 1}\left(1 - \left(\frac{P_j - P_i}{1 - P_i}\right)^{N - m_1}\right)\right\}.$$

Now, $m_1 \sim \text{binomial}(N, P_i)$ and from (2.2)

$$\mathbb{E}\left\{\frac{m_1}{N - m_1 + 1}\right\} = \frac{P_i}{1 - P_i}\left(1 - P_i^N\right).$$

Also,

$$\mathbb{E}\left\{\left(\frac{P_j - P_i}{1 - P_i}\right)^{N - m_1} \frac{m_1}{N - m_1 + 1}\right\} =$$

$$\sum_{m_1=0}^{N} \frac{m_1}{N - m_1 + 1} \binom{N}{m_1} P_i^{m_1} (1 - P_i)^{N - m_1} \left(\frac{P_j - P_i}{1 - P_i}\right)^{N - m_1} = \frac{P_i}{P_j - P_i}\left\{P_j^N - P_i^N\right\}.$$

Then,

$$\mathbb{E}\left\{\frac{m_1 m_2}{(N-m_1+1)(N-m_1-m_2+1)}\right\} =$$

$$\frac{P_j-P_i}{1-P_j}\left\{\frac{P_i}{1-P_i}-\frac{P_i^N}{1-P_i}-\frac{P_i}{P_j-P_i}P_j^N+\frac{P_i}{P_j-P_i}P_i^N\right\} = \frac{P_i(P_j-P_i)}{(1-P_i)(1-P_j)}+o\left(\frac{1}{N}\right),$$

and thus,

$$\mathbb{E}(c_i c_j) = \frac{p_i p_j}{(1-P_i)(1-P_j)} + o\left(\frac{1}{N}\right). \tag{2.9}$$

Therefore, for sufficiently small Δ,

$$\mathbb{E}\left\{Y_i Y_j\right\} = \mathbb{E}\left\{\frac{1}{\Delta^2}c_i c_j\right\} = \lambda(x_i)\lambda(x_j) + o\left(\frac{1}{N}\right).$$

Next we give a brief introduction to local linear fitting and together with the empirical estimate of this section, we use it to obtain an explicit expression of a kernel-based estimate for the hazard rate function.

2.3 Local linear fitting.

In nonparametric settings, the local linear fit method was initiated by Stone [42] and Cleveland [8]. Fan [12] refined this method by using smooth kernels weights to the least squares. For a summary of the available literature on the topic, see the references in section 2.3, of [13]. The essence of the method is to approximate the regression function, m, locally by

$$\lambda(x_0) \equiv m(x_0) \simeq m(x) + m'(x)(x_0 - x)$$

for x_0 in a neighborhood of x. This means that the hazard rate $\lambda(x_0)$ is modelled locally by the simple linear regression model

$$Y_i = \beta_0 + \beta_1(x_i - x) + \sigma(x_i)\varepsilon_i, \quad i = 1, \ldots, n$$

where $\beta_0 = \lambda(x)$ and $\beta_1 = \lambda'(x)$ are to be estimated. That is, we now regard hazard rate estimation as heteroscedastic nonparametric regression problem. The estimates of β_0 and β_1 will now be obtained by fitting a line locally using a weighted least squares method. Thus the estimates of β_0 and β_1 will result from the minimization of

$$\sum_{i=1}^{n}\left\{Y_i - \beta_0 - \beta_1(x_i - x)\right\}^2 K_h\left(x_i - x\right), \tag{2.10}$$

with respect to β_0 and β_1, where

$$K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$$

is a kernel function which assigns weight to each point and h is a bandwidth which controls the size of the local neighborhood. A theoretical justification of this particular choice of weights can be found in Wand and Jones [47]. The use of a kernel function K_h scaled by a bandwidth h as weight, ensures that observations closer to x will have more influence on the regression estimate at x than those further away. Moreover, the amount of the relative influence of the observations on the regression estimate is controlled by the bandwidth h so there is an analogy with the usual kernel estimators. When h is small then the estimate tends to take the shape of the data and when h is large the estimate tends to be the least squares line based on all data at once.

Of course, since $\lambda(x) = \beta_0$, we only need to find the estimate $\hat{\beta}_0$ of β_0 and is obtained by solving the minimization problem (2.10). Exact solution of this problem is given in appendix A.1, in subsection A.1.1. There, we show that

$$\hat{\beta}_0(x) =$$
$$\frac{\sum_{i=1}^{n} Y_i K_h(x_i - x)(x_i - x) \sum_{i=1}^{n} K_h(x_i - x)(x_i - x) - \sum_{i=1}^{n} Y_i K_h(x_i - x) \sum_{i=1}^{n} K_h(x_i - x)(x_i - x)^2}{\left(\sum_{i=1}^{n} Y_i K_h(x_i - x)(x_i - x)\right)^2 - \sum_{i=1}^{n} Y_i K_h(x_i - x)(x_i - x) \sum_{i=1}^{n} K_h(x_i - x)(x_i - x)^2}.$$

Setting

$$T_{n,l}(x) = \sum_{i=1}^{n} Y_i K\left(\frac{x_i - x}{h}\right)(x_i - x)^l, \quad l = 0, 1$$
$$\text{and} \quad S_{n,l}(x) = \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)(x_i - x)^l, \quad l = 0, 1, 2$$

we write this estimate more conveniently as

$$\hat{\beta}_0(x) \equiv \hat{\lambda}_L(x) = \frac{T_{n,1}(x)S_{n,1}(x) - T_{n,0}(x)S_{n,2}(x)}{S_{n,1}(x)S_{n,1}(x) - S_{n,0}(x)S_{n,2}(x)}.$$

We mention here that a variance stabilizing transformation such as Anscombe's transformation (see [13], pp. 48) may be needed in practical applications of the estimator. As our intention is simply to demonstrate the method and show that it is worth extending it to the hazard rate case, we defer the study of the transformed version of $\hat{\lambda}_L$ for future work.

In the next section we derive formulas for the asymptotic mean and variance of $\hat{\lambda}_L$.

2.4 Asymptotic properties.

For the boundary behavior of the estimator we concentrate on the interval $[0, h)$ as treatment of the right boundary is similar. The following assumptions on λ and the kernel are necessary for the study of the asymptotic properties of $\hat{\lambda}_L$.

A.1 λ has two derivatives and λ'' is bounded and uniformly continuous in the right neighborhood of zero (in a neighborhood of x) when estimation takes place at a boundary point (interior).

A.2

$$\int K^2 < +\infty \text{ and } \int |u^2 K| < +\infty$$

A.3

$$\int K = 1 \text{ and } \int uK = 0$$

A.4 For $l = 0, 1, 2$, $K^{(l)}$ is bounded and absolutely integrable with finite second moments.

The asymptotic properties of $\hat{\lambda}_L$ are summarized in the following theorem

Theorem 2.4.1. *Suppose that conditions A.1-A.4 hold and that $h \to 0$, $Nh \to +\infty$, $\Delta/h \to 0$. Then, if x is away from the boundary,*

$$\begin{aligned}
\mathbb{E}\hat{\lambda}_L(x) - \lambda(x) &= \frac{h^2}{2}\lambda''(x) \int u^2 K(u)\, du + o\left(h^2\right), \\
\mathbb{V}\mathrm{ar}\left\{\hat{\lambda}_L(x)\right\} &= \frac{1}{Nh}\frac{\lambda(x)}{1-F(x)} \int K^2(u)\, du + o\left(\frac{1}{Nh}\right).
\end{aligned}$$

If x is a boundary point, that is, $x = ph$, $p \geq 0$, then,

$$\begin{aligned}
\mathbb{E}\hat{\lambda}_L(x) - \lambda(x) &= \frac{h^2}{2}\lambda''(x) \int_{-p}^{+\infty} u^2 K_{(p)}(u)\, du + o\left(h^2\right), \\
\mathbb{V}\mathrm{ar}\left\{\hat{\lambda}_L(x)\right\} &= \frac{1}{Nh}\frac{\lambda(x)}{1-F(x)} \int_{-p}^{+\infty} K^2_{(p)}(u)\, du + o\left(\frac{1}{Nh}\right)
\end{aligned}$$

where

$$K_{(p)}(u) = \frac{s_{2,p} - s_{1,p}u}{s_{2,p}s_{0,p} - s^2_{1,p}} K(u)$$

and

$$s_{j,p} = \int_{-p}^{+\infty} u^j K(u)\, du, \quad j = 0, 1, 2.$$

Proof. We will prove the result for the boundary case since for the interior, the result follows from the boundary case by taking p sufficiently large and by condition A.3. Now, write $\hat{\lambda}_L(x)$ as

$$\begin{aligned}
\hat{\lambda}_L(x) &= \sum_{i=1}^{n} \frac{S_{n,2}(x) - S_{n,1}(x)(x_i - x)}{S_{n,2}(x)S_{n,0}(x) - S^2_{n,1}(x)} Y_i K\left(\frac{x_i - x}{h}\right) \\
&= \sum_{i=1}^{n} K_{N,x}\left(\frac{x_i - x}{h}\right) c_i
\end{aligned}$$

where

$$K_{N,x}(u) = \frac{S_{n,2}(x) - S_{n,1}(x)hu}{\Delta\left\{S_{n,2}(x)S_{n,0}(x) - S^2_{n,1}(x)\right\}} K(u).$$

Now, applying lemma 2.4.1, which is given below, with

$$t_i = \frac{x_i - x}{h}, \quad B = \frac{\Delta}{h}, \quad G(u) = u^l K(u) I_{[-p,\, +\infty)}(u)$$

gives

$$\begin{aligned} S_{n,l}(x) &= \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)(x_i - x)^l = \frac{1}{B}\sum_{i=1}^{n} K(t_i) t_i^l h^l B \\ &= \frac{h^l}{B}\sum_{i=1}^{n} K(t_i) t_i^l B = \frac{h^{l+1}}{\Delta}\left\{\int_{-p}^{+\infty} s^l K(s)\, ds + O(B^2)\right\}. \end{aligned}$$

By assumption $O(B^2)$ is $o(1)$ and so

$$S_{n,l}(x) = \frac{h^{l+1}}{\Delta} s_{l,p}(1 + o(1)), \quad l = 0, 1, 2. \tag{2.11}$$

Then it is easy to see that

$$\begin{aligned} K_{N,x}(u) &= \frac{\frac{h^3}{\Delta} s_{2,p} - \frac{h^2}{\Delta} s_{1,p} h u}{\Delta\left(\frac{h^4}{\Delta^2} s_{2,p} s_{0,p} - \frac{h^4}{\Delta^2} s_{1,p}^2\right)} K(u)(1 + o(1)) \\ &= \frac{1}{h}\frac{s_{2,p} - s_{1,p} u}{s_{2,p} s_{0,p} - s_{1,p}^2} K(u)(1 + o(1)) = \frac{1}{h} K_{(p)}(u)(1 + o(1)). \end{aligned} \tag{2.12}$$

From (2.12) and the first part of (2.8),

$$\mathbb{E}\left\{\hat{\lambda}_L(x)\right\} = \sum_{i=1}^{n} K_{N,x}\left(\frac{x_i - x}{h}\right)\mathbb{E}c_i = \frac{1}{h}\sum_{i=1}^{n} K_{(p)}\left(\frac{x_i - x}{h}\right)\Delta\lambda(x_i)(1 + o(1)).$$

Applying again lemma 2.4.1 with

$$t_i = x_i, \quad B = \Delta, \quad G(u) = K_{(0)}\left(\frac{u - ph}{h}\right)\lambda(u) I_{[0,\, +\infty)}(u), \quad K_{(0)} = \frac{s_{2,0} - s_{1,0} u}{s_{2,0} s_{0,0} - s_{1,0}^2} K(u)$$

we find,

$$\frac{1}{h}\sum_{i=1}^{n} K_{(p)}\left(\frac{x_i - x}{h}\right)\Delta\lambda(x_i)(1 + o(1)) = \frac{1}{h}\int_{0}^{+\infty} K_{(p)}\left(\frac{s - x}{h}\right)\lambda(s)\, ds(1 + o(1))$$

Setting $s - x = th$ we find

$$\mathbb{E}\left\{\hat{\lambda}_L(x)\right\} = \int_{-p}^{+\infty} K_{(p)}(t)\,\lambda(x + ht)\, dt(1 + o(1)).$$

Expanding $\lambda(x + ht)$ in Taylor series around x and noting that

$$\int K_{(p)}(t)\, dt = 1 \text{ and } \int t K_{(p)}(t)\, dt = 0$$

we finally get

$$\mathbb{E}\left\{\hat{\lambda}_L(x)\right\} = \lambda(x) + \frac{h^2}{2}\lambda''(x)\int_{-p}^{+\infty} t^2 K_{(p)}(t)\,dt + o(h^2).$$

As for the variance we have,

$$\mathbb{V}\mathrm{ar}\left\{\hat{\lambda}_L(x)\right\} = \mathbb{E}\left\{\hat{\lambda}_L^2(x)\right\} - \left(\mathbb{E}\left\{\hat{\lambda}_L(x)\right\}\right)^2. \tag{2.13}$$

Using (2.12) in the second step below,

$$\begin{aligned}\mathbb{E}\left\{\hat{\lambda}_L^2(x)\right\} &= \sum_{i=1}^{n}\sum_{j=1}^{n} K_{N,x}\left(\frac{x_i - x}{h}\right) K_{N,x}\left(\frac{x_j - x}{h}\right)\mathbb{E}\left(c_i c_j\right)\\ &= \frac{1}{h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_{(p)}\left(\frac{x_i - x}{h}\right) K_{(p)}\left(\frac{x_j - x}{h}\right)\mathbb{E}\left(c_i c_j\right)(1 + o(1)). \end{aligned}\tag{2.14}$$

By separating squared and cross product terms in (2.14),

$$\mathbb{E}\left\{\hat{\lambda}_L^2(x)\right\} = \frac{1}{h^2}\sum_{i=1}^{n} K_{(p)}^2\left(\frac{x_i - x}{h}\right) E(c_i^2)(1 + o(1))$$

$$+\frac{1}{h^2}\sum\sum_{i\neq j} K_{(p)}\left(\frac{x_i - x}{h}\right) K_{(p)}\left(\frac{x_j - x}{h}\right)\mathbb{E}\left(c_i c_j\right)(1 + o(1)).$$

By (2.8) and (2.9),

$$\begin{aligned}\mathbb{E}\left\{\hat{\lambda}_L^2(x)\right\} &= \frac{1}{h^2}\sum_{i=1}^{n} K_{(p)}^2\left(\frac{x_i - x}{h}\right)\left\{\frac{\Delta\lambda(x_i)}{N(1 - F(x_i))} + \Delta^2\lambda^2(x_i)\right\}(1 + o(1))\\ &+ \frac{1}{h^2}\sum\sum_{i\neq j} K_{(p)}\left(\frac{x_i - x}{h}\right) K_{(p)}\left(\frac{x_j - x}{h}\right)\left\{\Delta^2\lambda(x_i)\lambda(x_j)\right\}(1 + o(1)).\end{aligned}$$

Rearranging,

$$\begin{aligned}\mathbb{E}\left\{\hat{\lambda}_L^2(x)\right\} &= \frac{1}{h^2}\sum_{i=1}^{n} K_{(p)}^2\left(\frac{x_i - x}{h}\right)\frac{\Delta\lambda(x_i)}{N(1 - F(x_i))}(1 + o(1))\\ &+ \frac{1}{h^2}\sum_{i=1}^{n} K_{(p)}^2\left(\frac{x_i - x}{h}\right)\Delta^2\lambda^2(x_i)(1 + o(1))\\ &+ \frac{1}{h^2}\sum\sum_{i\neq j} K_{(p)}\left(\frac{x_i - x}{h}\right) K_{(p)}\left(\frac{x_j - x}{h}\right)\left\{\Delta^2\lambda(x_i)\lambda(x_j)\right\}(1 + o(1)).\end{aligned}$$

Writing the last two sums as one term,

$$\begin{aligned}\mathbb{E}\left\{\hat{\lambda}_L^2(x)\right\} &= \frac{1}{h^2}\sum_{i=1}^{n} K_{(p)}^2\left(\frac{x_i - x}{h}\right)\frac{\Delta\lambda(x_i)}{N(1 - F(x_i))}(1 + o(1))\\ &+ \frac{1}{h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_{(p)}\left(\frac{x_i - x}{h}\right) K_{(p)}\left(\frac{x_j - x}{h}\right)\left\{\Delta^2\lambda(x_i)\lambda(x_j)\right\}(1 + o(1)).\end{aligned}$$

Thus,

$$\mathbb{E}\left\{\hat{\lambda}_L^2(x)\right\} = \frac{1}{h^2}\sum_{i=1}^{n} K_{(p)}^2\left(\frac{x_i - x}{h}\right)\frac{\Delta\lambda(x_i)}{N(1-F(x_i))}(1+o(1)) + \left\{\mathbb{E}\left(\hat{\lambda}_L(x)\right)\right\}^2.$$

Substituting back to (2.13) and using lemma 2.4.1 in the second step below,

$$\begin{aligned}\mathbb{V}\text{ar}\left\{\hat{\lambda}_L(x)\right\} &= \frac{1}{h^2}\sum_{i=1}^{n} K_{(p)}^2\left(\frac{x_i - x}{h}\right)\frac{\lambda(x_i)}{N(1-F(x_i))}(1+o(1)) \\ &= \frac{1}{Nh^2}\int_0^{+\infty} K_{(p)}^2\left(\frac{x-u}{h}\right)\frac{\lambda(u)}{1-F(u)}\,du(1+o(1)).\end{aligned}$$

Setting $x - u = ht$ and expanding $\lambda(x+ht)/(1-F(x+ht))$ in Taylor series around x we find

$$\mathbb{V}\text{ar}\left\{\hat{\lambda}_L(x)\right\} = \frac{1}{Nh}\frac{\lambda(x)}{1-F(x)}\int_{-p}^{+\infty} K_{(p)}^2(t)\,du + o\left(\frac{1}{Nh}\right). \quad \blacksquare$$

Lemma 2.4.1. *Let G be a real valued function with domain $D(G)$ and t_i, $i = 1, 2, \ldots, n$ is a set of equally spaced points on $D(G)$ with grid width B. If G is twice differentiable and its second derivative is integrable, then*

$$\left|\sum_{i=1}^{n} G(t_i)B - \int G(s)\,ds\right| \leq \frac{B^2}{4}\int |G''(t)|\,dt.$$

Remark 2.1. Essentially the lemma says that if the grid width B is small, then

$$\sum_{i=1}^{n} G(t_i)B \simeq \int G(s)\,ds,$$

i.e. the usual Riemman sum approximation, with more precise bounds.

The proof of the lemma is given in the appendix in subsection A.1.2.

2.5 Bandwidth selection.

The performance of the local linear hazard rate estimator depends on the bandwidth h which controls the amount of smoothing applied to local neighborhoods. Therefore, it is important to find good data-based procedures for selection of bandwidth.

For this, the global error criterion we use is integrated square error. That is, integrated square error of $\hat{\lambda}_L$ which uses bandwidth h is,

$$\text{ISE}\left(\hat{\lambda}_L(x); h\right) = \int \left(\lambda(x) - \hat{\lambda}_L(x)\right)^2 w(x)\,dx,$$

where $w(x)$ is an appropriately defined weight function. For the discussion of this section we take the weight function w to be bounded and supported on $[0, T]$ where $T = \sup\{x : F(x) < 1 - \varepsilon\}$ for a small $\varepsilon > 0$.

In subsection 2.5.1 we discuss a plug-in rule. We first find the ideal bandwidth which minimizes the above error over all possible data sets (that is, a bandwidth which minimizes mean ISE). As expected such a choice involves unknown functions and hence, for practically useful bandwidth choice, we provide suitable estimators for the unknown functions in the ideal bandwidth. In subsection 2.5.2 we develop the least squares cross-validation bandwidth. As opposed to the plug-in rule, this method chooses a bandwidth which is best for the given data set.

2.5.1 Simple plug-in rule.

Inspired by the outstanding performance of the Sheather and Jones bandwidth selector in [38] we attempt to bring in the root finding plug-in idea to the hazard rate case. The Sheather and Jones bandwidth selector is motivated by the explicit expression of the asymptotic optimal bandwidth for kernel estimators. Below, we show that a similar representation holds also for the hazard rate case and thus it is natural to consider the plug-in method.

From the bias and variance expressions of theorem 2.4.1, the MSE is

$$\text{MSE}(\hat{\lambda}_L(x)) = \frac{1}{4}h^4\mu_2^2(K)\lambda''(x)^2 + \frac{1}{Nh}\frac{\lambda(x)}{1-F(x)}\int K^2(x)\,dx + o\left(\frac{1}{Nh}\right) + O\left(h^4\right).$$

Let T be as defined above and suppose that conditions $A.1 - A.4$ hold. Integrating, we find the MISE to be

$$\text{MISE}(\hat{\lambda}_L(x)) = \frac{1}{4}h^4\mu_2^2(K)R(\lambda'') + \frac{1}{Nh}R(K)\int \frac{\lambda(x)w(x)}{1-F(x)}\,dz + o\left(\frac{1}{Nh}\right) + O\left(h^4\right),$$

where

$$R(g(x)) = \int g^2(x)w(x)\,dx.$$

With the additional assumption that the kernel is Hölder continuous, and as $N \to +\infty$, $h \to 0$ and $Nh \to +\infty$, the MISE can be well approximated by the AMISE. In this case it is

$$\text{AMISE}(\hat{\lambda}_L(x)) = \frac{1}{4}h^4\mu_2^2(K)R(\lambda'') + \frac{1}{Nh}R(K)\int \frac{\lambda(x)w(x)}{1-F(x)}\,dx. \tag{2.15}$$

Minimizing this quantity with respect to h gives AMISE-optimal bandwidth. Differentiating (2.15) with respect to h yields

$$h^3\mu_2^2(K)R(\lambda'') - \frac{1}{Nh^2}R(K)\int \frac{\lambda(x)w(x)}{1-F(x)}\,dx = 0$$

and therefore

$$h_{AMISE} = \left\{\frac{R(K)M}{N\mu_2^2(K)R(\lambda'')}\right\}^{\frac{1}{5}} \tag{2.16}$$

with M being

$$M = \int_0^T \frac{\lambda(x)}{1-F(x)}\,dx = \int_0^T \frac{f(x)}{(1-F(x))^2}\,dx = \frac{1}{1-F(T)} - 1$$

The quantities $R(\lambda'')$ and M involve the unknown hazard rate and the true cdf F respectively and therefore they have to be estimated. However, it can be proved that if good enough estimators of the unknown quantities can be found, then the bandwidth resulting from the plug-in method, say $\hat{h}$, *converges* to the optimal, for example

$$\hat{h} \xrightarrow{p} h_{AMISE}.$$

Interesting issues here are the choice of estimator for $R(\lambda'')$ (e.g. usual kernel estimator, local linear etc.) and correspondingly, rate of convergence of $\hat{h}$ to h_{AMISE}. In the present work, with practical implementation being the primary objective, we give a simple plug-in formulation below and the detail analysis of these issues will be carried out in future work.

An obvious estimator for M is

$$\hat{M} = \frac{1}{1 - F_N(T)} - 1,$$

where F_N is the empirical cdf of the sample. As about the functional $R(\lambda'')$, its general form is

$$R\left(\lambda^{(s)}\right) = \int \lambda^{(s)}(x)^2 w(x)\, dx.$$

In appendix A.1.3 we prove that under sufficient smoothness assumptions on λ

$$R\left(\lambda^{(s)}(x)\right) = (-1)^s \int \lambda^{(2s)}(x)\lambda(x)w(x)\, dx.$$

Set

$$\psi_r = \int \lambda^{(r)}(x)\lambda(x)w(x)\, dx, \quad \text{for } r \text{ even}$$

and observe that

$$\mathbb{E}\left\{\frac{\lambda^{(r)}(X)}{1 - F(X)} w(X)\right\} = \int \frac{\lambda^{(r)}(x)}{1 - F(x)} f(x)w(x)\, dx = \psi_r. \tag{2.17}$$

Estimating ψ_r by a kernel estimate of

$$\frac{\lambda^{(r)}(x)}{1 - F(x)} w(x),$$

say $\hat{\psi}_r$, has the disadvantage that this estimate will depend on the bandwidth of the kernel, say g, so the method will not be fully automatic. One way around this is to use the AMISE-optimal bandwidth similar to (2.16) but the resulting bandwidth will depend on the unknown functional ψ_{r+2}. Moreover an estimate of ψ_{r+2} will depend on ψ_{r+4}, i.e. we have a recursive estimation of $\psi_{r+i}, i = 2, 4, 6, \ldots$ without end. A common way to overcome this problem is to calculate the AMISE optimal bandwidth for the kernel estimate of ψ_{r+2} functional by estimating ψ_{r+4} with a reference to some distribution. Once ψ_{r+2} is found then we can plug it in a formula similar to (2.16) to find the optimal bandwidth for $\hat{\psi}_r$. An issue that still needs to be answered is the number of stages, l, (or the number of functionals) that we need to estimate to find the optimal bandwidth for $\hat{\psi}_r$. For the density case Wand and Jones [47] suggest $l = 2$.

Estimating ψ_r with reference to some distribution is essentially computation of ψ_r with λ taken from a standard distribution. The weibull distribution is a frequently met one when it comes to hazard rate models, thus it makes a sensible choice. We assume that the scale parameter of the distribution is β and the index κ. Then (see appendix A.1.4),

$$\psi_r = \frac{\kappa^{2\beta}\beta^2(\beta-1)\dots(\beta-r)}{2\beta-r-1}b^{2\beta-r-1}.$$

Since the issue of bandwidth selection for ψ_{r+2} is resolved we turn our attention to the kernel estimator of ψ_r. From (2.17), an intuitive choice as an estimate of ψ_r is

$$\hat{\psi}_r^*(g) = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\lambda}^{(r)}(X_i;g)}{1-F_N(X_i)}w(X_i).$$

with g representing the pilot bandwidth. Denote with L a kernel, possibly different than K. Set

$$L_g(\cdot) = \frac{1}{g}L\left(\frac{\cdot}{g}\right).$$

Then, $\hat{\psi}_r^*(g)$ can be written as

$$\begin{aligned}\hat{\psi}_r^*(g) &= \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{L_g^{(r)}\left(X_{(i)}-X_{(j)}\right)}{(N-j+1)(N-i+1)}w(X_{(i)})w(X_{(j)}) \\ &= \sum_{i=1}^{N}\frac{L_g^{(r)}(0)}{N-i+1}w(0) + \sum\sum_{i\neq j}\frac{L_g^{(r)}\left(X_{(i)}-X_{(j)}\right)}{(N-j+1)(N-i+1)}w(X_{(i)})w(X_{(j)}).\end{aligned}$$

The first term on the RHS of the above equation is a constant and hence may be thought of adding a type of bias to the estimator. This motivates the estimator

$$\hat{\psi}_r(g) = \sum\sum_{i\neq j}\frac{L_g^{(r)}\left(X_{(i)}-X_{(j)}\right)}{(N-j+1)(N-i+1)}w(X_{(i)})w(X_{(j)}).$$

The asymptotic properties of this estimator are crucial for its performance. The algebra involved in studying the precise asymptotic properties of $\hat{\psi}_r(g)$ is tedious. That can be seen in section A.1.5, lemma A.1.1 where we calculate the mean of estimator $\hat{\psi}_r(g)$. Note also that with a similar calculation for the variance we can obtain an upper bound. However, in order to get insight into the asymptotic MSE of $\hat{\psi}_r^*(g)$ we can use the fact that

$$\begin{aligned}\frac{L_g^{(r)}(X_i-X_j)}{(1-F_n(X_i))(1-F_n(X_j))} &= \frac{L_g^{(r)}(X_i-X_j)}{(1-F(X_i))(1-F(X_j))} \\ &\quad+ \frac{L_g^{(r)}(X_i-X_j)}{(1-F_n(X_i))(1-F_n(X_j))} - \frac{L_g^{(r)}(X_i-X_j)}{(1-F(X_i))(1-F(X_j))} \\ &= \frac{L_g^{(r)}(X_i-X_j)}{(1-F(X_i))(1-F(X_j))} + o_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

and consider the asymptotically equivalent estimator

$$\hat{\psi}'_r(g) = \sum\sum_{i \neq j} \frac{L_g^{(r)}\left(X_i - X_j\right)}{(1 - F(X_i))(1 - F(X_j))} w(X_i) w(X_j).$$

We summarize the asymptotic properties of $\hat{\psi}'_r(g)$ them in the following theorem.

Theorem 2.5.1. *Assume that L is a symmetric kernel of order k, $k = 2, 4, \dots$, possessing r derivatives. Suppose also that λ has $p > k$ continuous derivatives and that $g = g_N$ is a positive sequence of bandwidths satisfying*

$$\lim_{N \to +\infty} g = 0, \quad \lim_{N \to +\infty} N g^{r+1} = +\infty$$

Then, with $\mu_k(L) = \int (u^k/k!) L(u)\, du$

$$\begin{aligned}
\mathbb{E}\hat{\psi}'_r(g) &= \psi_r + \frac{\mu_k(L) g^k}{k!} \psi_{r+k} + O\left(g^{k+2}\right), \\
\mathbb{V}\mathrm{ar}\left(\hat{\psi}'_r(g)\right) &= \frac{\psi_0}{n^2 g^{2r+1}} R\left(L^{(r)}\right) \int \left\{\frac{\lambda(z)}{1 - F(z)}\right\}^2 dz - \left(\mathbb{E}\hat{\psi}'_r(g)\right)^2 + o\left(\frac{1}{g^{2r+1}}\right).
\end{aligned}$$

Proof. From section A.1.5, lemma A.1.1,

$$\mathbb{E}\hat{\psi}'_r(g) = \iint L_g^{(r)}(x - y) \lambda(x) \lambda(y)\, dx\, dy.$$

Using integration by parts gives

$$\mathbb{E}\hat{\psi}'_r(g) = \iint L_g(x - y) \lambda(x) \lambda^{(r)}(y)\, dx\, dy.$$

Applying the change of variable $x - y = gu$ and expanding $\lambda^{(r)}$ in Taylor series around y yields

$$\begin{aligned}
\mathbb{E}\hat{\psi}'_r(g) &= \iint L(u) \lambda(y + gu) \lambda^{(r)}(y)\, du\, dy \\
&= \iint L(u) \lambda^{(r)}(y) \left\{\sum_{l=0}^{k} \frac{(ug)^l}{l!} \lambda^{(l)}(y) + O\left(g^{k+1}\right)\right\} du\, dy \\
&= \psi_r + \frac{\mu_k(L) g^k}{k!} \psi_{r+k} + O\left(g^{k+2}\right).
\end{aligned}$$

As about the variance, from section A.1.5, lemma A.1.2,

$$\mathbb{V}\mathrm{ar}\left(\hat{\psi}'_r(g)\right) = \iint L_g^{(r)}(x - y)^2 \frac{\lambda(x)\lambda(y)}{(1 - F(x))(1 - F(y))}\, dx\, dy - \left(\mathbb{E}\hat{\psi}'_r\right)^2.$$

Let

$$s(x) = \frac{\lambda(x)}{1 - F(x)}.$$

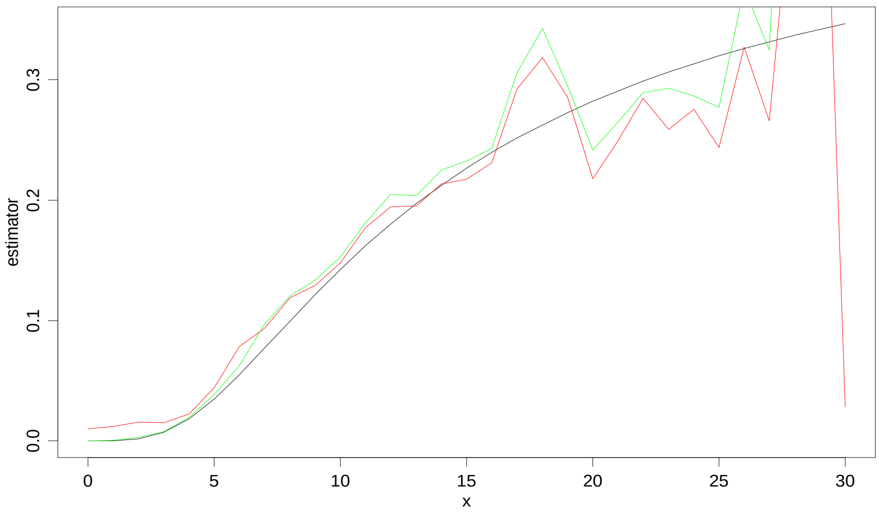


Figure 2.1: Comparison of $\hat{\lambda}_L$ (green line) with estimator $\hat{\lambda}_2$ (red line). The black line is the true hazard rate of the χ^2_{12} distribution.

Now, set $x - y = gu$. Then, by expanding $s(y + gu)$ in Taylor series around y

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\hat{\psi}'_r(g)\right) &= \frac{1}{n^2 g^{2r+1}} \iint L_g^{(r)}(u)^2 s(y+gu)s(y)\,du\,dy - \left(\mathbb{E}\hat{\psi}'_r(g)\right)^2 \\ &= \frac{1}{n^2 g^{2r+1}} R\left(L^{(r)}\right) \int s^2(z)\,dz - \left(\mathbb{E}\hat{\psi}'_r(g)\right)^2 + o\left(\frac{1}{g^{2r+1}}\right). \quad \blacksquare\end{aligned}$$

In figure 2.1 we illustrate the performance of the local linear fit estimator by comparing it with estimator $\hat{\lambda}_2$ (defined in the introduction). We discretize a sample of 1000 values from the χ^2_{12} distribution to 100 values over the interval $[0, 30]$. Both estimators use the biweight kernel,

$$K(t) = \begin{cases} \frac{15}{16}(1-t^2)^2 & |t| < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The initial sample is used to plot estimator $\hat{\lambda}_2$, taking as bandwidth $h = 1.26$ after using Silverman's default bandwidth method ([39]). The resulting estimator is the red line. The green line is estimator $\hat{\lambda}_L$ and uses bandwidth $h = 1.5$ selected by the plug-in rule. Obviously, $\hat{\lambda}_L$ reduces the bias at the left boundary compared to $\hat{\lambda}_2$. Between $x = 7$ and $x = 20$, $\hat{\lambda}_2$ appears to be more accurate, although $\hat{\lambda}_L$ is quite acceptable as an estimate of the curve. In general, away from the boundary, the performance of both estimators is similar and this is something expected since $\hat{\lambda}_L$ is essentially the same as a conventional kernel estimator in the interior. Next we discuss the least squares cross validation method for choosing the bandwidth for $\hat{\lambda}_L$.

2.5.2 Least squares cross validation.

The Integrated Squared Error for $\hat{\lambda}_L$ is

$$\text{ISE}\left(h; \hat{\lambda}_L\right) = \int \left(\hat{\lambda}_L(x) - \lambda(x)\right)^2 w(x)\, dx.$$

Minimizing the ISE with respect to h defines another criterion for bandwidth choice. Expand the integrand to get

$$\text{ISE}\big(h; \hat{\lambda}_L\big) = \int \hat{\lambda}_L^2(x) w(x)\, dx - 2\int \hat{\lambda}_L(x)\lambda(x) w(x)\, dx + \int \lambda^2(x) w(x)\, dx.$$

Since the third term in the above equation does not depend on h it is equivalent to consider only minimization of the first two. Let

$$S\big(h; \hat{\lambda}_L\big) = \int \hat{\lambda}_L^2(x) w(x)\, dx - 2\int \hat{\lambda}_L(x)\lambda(x) w(x)\, dx. \tag{2.18}$$

The problem is that the second term includes the unknown hazard rate. Consider the leave-one-out version of $\hat{\lambda}_L$ given by

$$\hat{\lambda}_L^{-i}(x_i) = \sum_{j\neq i} K_{N,x_j}\left(\frac{x_i - x_j}{h}\right) c_j w(X_i)$$

and notice that

$$\begin{aligned}
\mathbb{E}\left\{\sum_{i=1}^{n} c_i \hat{\lambda}_L^{-i}(x_i) w(x_i)\right\} &= \sum_{i=1}^{n}\sum_{j\neq i} K_{N,x_j}\left(\frac{x_i - x_j}{h}\right)\mathbb{E}(c_i c_j) w(X_i) \\
&= \sum_{i=1}^{n}\sum_{i\neq j} K_{N,x_j}\left(\frac{x_i - x_j}{h}\right)\Delta^2 \lambda(x_i)\lambda(x_j) w(X_i).
\end{aligned}$$

Using lemma 2.4.1,

$$\mathbb{E}\left\{\sum_{i=1}^{n} c_i \hat{\lambda}_L^{-i}(x_i) w(X_i)\right\} = \int \sum_{i\neq j} K_{N,x_j}\left(\frac{x - x_j}{h}\right)\Delta\lambda(x)\lambda(x_j) w(x)\, dx$$

and therefore,

$$\mathbb{E}\left\{\sum_{i=1}^{n} c_i \hat{\lambda}_L^{-i}(x_i) w(X_i)\right\} = \int \mathbb{E}\left\{\hat{\lambda}_L^{-i}(x)\right\}\lambda(x) w(x)\, dx. \tag{2.19}$$

As can be seen for the bias expression of $\hat{\lambda}_L$ in theorem 2.4.1, the mean of this estimator depends on the kernel and the bandwidth and not on the sample size. Therefore

$$\mathbb{E}\left\{\hat{\lambda}_L^{-i}(x)\right\} = \mathbb{E}\left\{\hat{\lambda}_L(x)\right\}.$$

Then, from (2.19)

$$\mathbb{E}\left\{\sum_{i=1}^{n} c_i \hat{\lambda}_L^{-i}(x_i) w(X_i)\right\} = \mathbb{E}\int \hat{\lambda}_L(x)\lambda(x) w(x)\, dx$$

and thus, we easily deduce that

$$\mathbb{E}S\big(h;\hat{\lambda}_L\big) = \mathbb{E}\left(\int \hat{\lambda}_L^2(x)w(x)\,dx - 2\int \hat{\lambda}_L^{-i}(x)\lambda(x)w(x)\,dx\right).$$

Therefore, we can replace the unknown hazard rate in (2.18) by,

$$\sum_{i=1}^{n} c_i\hat{\lambda}_L^{-i}(x_i)w(X_i).$$

Hence, we define the least squares cross validation criterion to be

$$CV(h) = \int \hat{\lambda}_L^2(x)w(x)\,dx - 2\sum_{i=1}^{n}\sum_{i\neq j}^{n} K_{N,x_j}\left(\frac{x_i - x_j}{h}\right) c_i c_j w(X_i). \tag{2.20}$$

This method has been studied by many authors for many kernel based estimates of the hazard rate (see [29], [28] and the references therein). However, in the present setting our cross validation function is analogous to the equivalent function in the bandwidth choice for nonparametric regression function, in particular, when local linear method is used. Treating this as a regression problem with n design points and response Y_i as defined earlier, minimization of the above cross validation criterion provides a reasonable bandwidth and our simulation had confirmed this. Asymptotic properties like asymptotic optimality and rate of convergence of the least squares cross validation bandwidth are very interesting and open problems. It should be noted that these asymptotic properties are to be studied not only by allowing n and hence sample size N to grow but would also involve at what rate design points increase in terms of N. All these issues will be analyzed in future work.

Chapter 3

VARIABLE BANDWIDTH FOR HAZARD RATE ESTIMATION

3.1 Introduction.

Asymptotically, in terms of precision in kernel density estimation, according to Hall, Hu and Marron [17] it is enough to consider only the rate of convergence of the bias. That is because most of the commonly used variations of the kernel estimator have variances of the same order, $\frac{1}{nh}$, so it's the bias we can reduce in order to get a better MSE.

The first bias reduction technique, proposed in the density estimation setting by Parzen [27] and Bartlett [4], involved the use of higher order kernels. With this method the rate of convergence of the bias improves from h^2 to a much faster rate of order h^{2r} when a $2r$th order kernel is used where r is an integer greater than 1. For further insight into higher order kernel density estimation see Wand and Marron [24]. It is important here to note that kernels of this type often lead to the undesirable feature of negative estimates in the tails. This, together with the fact that the hazard rate is a non negative function, makes the proposal of higher order kernels unattractive for nonparametric hazard rate estimation.

Again in density estimation settings, another approach to reduce the bias as proposed by Breiman *et al*. [5] and by Mack and Rosenblatt [23] is to use locally adjusted kernels, which allow the bandwidth to vary according to data points or according to the estimation location. Abramson [2] showed the improvement of the Breiman *et al*. approach, and he provided a law (often called the square root law) according to which the rate of convergence improves to h^4. In this case the resulting estimator behaves in several aspects as it was based on a fourth order kernel, where it is actually of order two and is nonnegative.

Hall and Marron [16] consider a variant of the estimator used by Abramson and adopting the square root law for the bandwidth showed that the bias rate is h^4, even if a conventional global kernel estimator is used as an estimate of the best bandwidth function. Most importantly their estimator integrates to one and it is nonnegative. For some further interesting results see Hall [14], Jones [19], Terrel and Scott [45] and Hall, Hu and Marron [17].

Motivated by these ideas we extend the concept of variable bandwidth to the case of the nonparametric failure rate estimation. We start in section 3.2 by extending estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ defined in (1.1) and (1.3) respectively, so that they make use of variable bandwidth, and in section 3.3 we prove that by doing so the bias decreases to $O(h^4)$. Interestingly the new estimators are non negative. Practically useful versions of the newly defined estimators are given in section 3.4, and in section 3.5 we summarize results on the performance of these estimators, i.e. that they attain the same rate of convergence as those defined in section 3.2. There, we also quantify the difference between the estimators defined in section 3.2 and their practically useful counterparts defined in section 3.4. Proofs of the main results of section 3.5 are given in section 3.6. Section 3.7 contains proofs of lemmas used to prove main results. Graphical performance on simulated as well as real life data is deferred until the second half of the next chapter.

3.2 Variable bandwidth estimators.

Recall the estimators defined in the first chapter,

$$\hat{\lambda}_1(x|h) = \frac{\hat{f}(x|h)}{1 - F_n(x)}$$
$$\hat{\lambda}_2(x|h) = \sum_{i=1}^{n} \frac{K_h\left(x - X_{(i)}\right)}{n - i + 1}.$$

The modification of $\hat{\lambda}_1$ to allow variable bandwidth already exists in the literature, see for example Silverman [39]. It involves replacing the numerator of $\hat{\lambda}_1$ with $\tilde{f}$, where $\tilde{f}$ uses bandwidth proportional to $f^{-\frac{1}{2}}$ and replacing F_n by

$$\hat{F}(x) = \int_{-\infty}^{x} \tilde{f}(u)\, du$$

i.e.,

$$\lambda^*_{n,1}(x) = \frac{\tilde{f}(x|h)}{1 - \hat{F}(x)}$$

where

$$\tilde{f}(x|h) = \frac{1}{nh} \sum_{i=1}^{n} f(X_i)^{\frac{1}{2}} K\left(\frac{x - X_i}{h} f(X_i)^{\frac{1}{2}}\right), \quad x \in \mathbb{R}.$$

With a slight deviation from the estimator in Silverman [39] we define

$$\tilde{\lambda}_{n,1}(x|h) = \frac{\tilde{f}(x|h)}{1 - F_n(x)}. \tag{3.1}$$

In the case of $\hat{\lambda}_2$, we modify this estimator by taking bandwidth proportional to $\lambda^{-\frac{1}{2}}$ to define a new estimator that makes use of variable bandwidth,

$$\begin{aligned}\tilde{\lambda}_{n,2}(x|h) &= \frac{1}{nh} \sum_{i=1}^{n} \frac{\lambda(X_i)^{\frac{1}{2}} K\left(\frac{x - X_i}{h} \lambda(X_i)^{\frac{1}{2}}\right)}{1 - F_n(X_i)} \\ &= \frac{1}{h} \sum_{i=1}^{n} \frac{\lambda(X_{(i)})^{\frac{1}{2}} K\left(\frac{x - X_{(i)}}{h} \lambda(X_{(i)})^{\frac{1}{2}}\right)}{n - i + 1}. \end{aligned} \tag{3.2}$$

This estimator is a natural extension of the density estimator defined by Hall and Marron in [16] to the case of failure rate estimation.

Note that the bandwidth law, $h \equiv h\lambda(x)^{-\frac{1}{2}}$, used in the definition of $\tilde{\lambda}_{n,2}$ involves the true hazard function, λ, just as the bandwidth law $h \equiv hf(x)^{-\frac{1}{2}}$ used in the variable bandwidth density estimator $\tilde{f}$ depends on the true density f. This makes the above estimators unfeasible in practice. However, in the next section, we show that these *ideal* variable bandwidth estimators have better bias and hence MSE (since their variances are of the same order) properties than their fixed bandwidth counterparts.

3.3 Mean square error analysis of the ideal variable bandwidth estimators.

First, in subsection 3.3.1, we consider the mean square error properties of $\tilde{\lambda}_{n,1}$. Due to the technical difficulties involved in the study of the expected value of $\tilde{\lambda}_{n,1}$, we follow the common method of approximating the estimator with its asymptotic version. Then, the mean square error analysis is very straightforward and hence we state the mean and variance formulae without proof. In subsection 3.3.2 we state and prove the mean square error properties of the most preferred estimator $\tilde{\lambda}_{n,2}$.

Throughout this chapter we assume the standard kernel conditions, that is, the kernel function $K : \mathbb{R} \to \mathbb{R}$, satisfies:

A.1
$$\int K(z)\, dz = 1.$$

A.2 K is nonnegative, symmetric and five times differentiable.

A.3 K vanishes outside a compact interval.

We note here that an implication of the last two conditions is that both the fourth moment of the kernel and its supremum are finite. Also, throughout this chapter we assume that the support of the hazard function λ is $[0, T]$ where $T = \sup\{x | 1 - F(x) > \varepsilon\}$, $\varepsilon > 0$.

3.3.1 Mean and variance of the $\tilde{\lambda}_{n,1}$ estimator.

Write $\tilde{\lambda}_{n,1}$ as

$$\tilde{\lambda}_{n,1}(x) = \frac{\tilde{f}(x|h)}{1 - F_n(x)} = \frac{\tilde{f}(x|h)}{1 - F(x) + F(x) - F_n(x)} = \frac{\tilde{f}(x|h)}{1 - F(x)} \frac{1}{1 + \frac{F(x) - F_n(x)}{1 - F(x)}}$$

and note that

$$\frac{1}{1 + \dfrac{F(x) - F_n(x)}{1 - F(x)}} = \sum_{i=0}^{\infty} (-1)^i \left(\frac{F(x) - F_n(x)}{1 - F(x)} \right)^i.$$

Since (see Serfling [37])

$$\sup_{-\infty<x<+\infty} |F_n(x) - F(x)| = O_p\left(n^{-\frac{1}{2}}\right),$$

we can write $\tilde{\lambda}_{n,1}$ as

$$\tilde{\lambda}_{n,1}(x|h) = \frac{\tilde{f}(x|h)}{1-F(x)}\left(1 + O_p\left(\frac{1}{\sqrt{n}}\right)\right).$$

Therefore, $\tilde{\lambda}_{n,1}$ can be approximated asymptotically by

$$\bar{\lambda}_{n,1}(x|h) = \frac{\tilde{f}(x|h)}{1-F(x)}$$

with negligible error. Now, let $g(x)$ be as in [16], p. 41, i.e.,

$$g(x) = \frac{24f'(x)^4 - 36f'(x)^2 f''(x)^2 f(x) + 6f''(x)^2 f^2(x) + 8f'(x)f'''(x)f^2(x) - f^{(4)}(x)f^3(x)}{24f^5(x)}$$

and set

$$r(x) = \frac{g(x)}{1-F(x)}.$$

Then we have the following theorem:

Theorem 3.3.1. *Assume conditions A1,A3 for the kernel. Assume also that f has four continuous derivatives and that f is bounded away from zero on $(0,T)$. Then*

$$\mathbb{E}\left\{\bar{\lambda}_{n,1}(x|h)\right\} = \lambda(x) + r(x)h^4 \int z^4 K(z)\,dz + o\left(h^4\right) \quad \textit{and}$$
$$\mathbb{V}\mathrm{ar}\left\{\bar{\lambda}_{n,1}(x|h)\right\} = \frac{1}{nh}\frac{f(x)^{\frac{3}{2}}}{(1-F(x))^2}\int K^2(z)\,dz + o\left(\frac{1}{nh}\right)$$

uniformly in $x \in (0,T)$ as $h \to 0$, $n \to +\infty$ and $nh \to +\infty$.

3.3.2 Mean and variance of the $\tilde{\lambda}_{n,2}$ estimator.

The asymptotic properties of $\tilde{\lambda}_{n,2}$ are summarized in the form of the following theorem

Theorem 3.3.2. *Under the assumptions A1,A3 for the kernel, and assuming that λ has four continuous derivatives and that it is bounded away from zero on $(0,T)$,*

$$\mathbb{E}\left\{\tilde{\lambda}_{n,2}(x|h)\right\} = \lambda(x) + g_1(x)h^4 \int z^4 K(z)\,dz + o(h^4)$$
$$\mathbb{V}\mathrm{ar}\left\{\tilde{\lambda}_{n,2}(x|h)\right\} = \frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)}\frac{1}{nh}\int K^2(z)\,dz + o\left(\frac{1}{nh}\right)$$

uniformly in $x \in (0,T)$ as $h \to 0$,$n \to +\infty$ and $nh \to +\infty$, with

$$g_1(x) = \frac{1}{24\lambda(x)^5}\Big(24\lambda'(x)^4 - 36\lambda'(x)^2\lambda''(x)^2\lambda(x) + 6\lambda''(x)^2\lambda^2(x) + 8\lambda'(x)\lambda'''(x)\lambda^2(x) - \lambda^{(4)}(x)\lambda^3(x)\Big).$$

Proof. First we obtain an integral expression for the bias, which is then evaluated by applying a transformation and finally its terms are expanded as power series of the transformed variable. To begin, from remark 1.2 we have

$$\mathbb{E}\left\{\hat{\lambda}(x|h)\right\} = \int \lambda(u) K_h(u-x)[1-F^n(u)]\, du.$$

Then,

$$\begin{aligned}
\mathbb{E}\left\{\tilde{\lambda}_{n,2}(x|h)\right\} &= \mathbb{E}\frac{1}{h}\sum_{i=1}^{n}\frac{\lambda(X_{(i)})^{\frac{1}{2}} K\left(\frac{x-X_{(i)}}{h}\lambda(X_{(i)})^{\frac{1}{2}}\right)}{n-i+1} \\
&= \frac{1}{h}\int [1-F^n(u)]\lambda(u)^{1/2} K\left(\frac{x-u}{h}\lambda(u)^{1/2}\right)\lambda(u)\, du \\
&= \frac{1}{h}\int [1-F^n(u)]\lambda(u)^{3/2} K\left(\frac{x-u}{h}\lambda(u)^{1/2}\right)\, du \\
&= \frac{1}{h}\int \lambda(u)^{\frac{3}{2}} K\left(\frac{x-u}{h}\lambda(u)^{\frac{1}{2}}\right)\, du - \frac{1}{h}\int \lambda(u)^{\frac{3}{2}} F^n(x) K\left(\frac{x-u}{h}\lambda(u)^{\frac{1}{2}}\right)\, du.
\end{aligned}$$

The second term tends to zero because $|F(x)| < 1$ so for large n we have that $|F(x)|^n = o(n^{-1})$. That is, the second term is $o(n^{-1})$. For the first term, applying the transformation

$$\frac{x-u}{h} = z$$

we get

$$x-u = hz \Leftrightarrow u = x-hz \Leftrightarrow du = -hdz.$$

Then

$$\begin{aligned}
\mathbb{E}\left\{\tilde{\lambda}_{n,2}(x|h)\right\} &= \int \lambda(x-hz)^{3/2} K\left(z\lambda(x-hz)^{1/2}\right) dz + o\left(\frac{1}{n}\right) \\
&= \int \lambda(x)^{\frac{3}{2}}\frac{\lambda(x-hz)^{\frac{3}{2}}}{\lambda(x)^{\frac{3}{2}}} K\left(z\lambda(x)^{\frac{1}{2}}\frac{\lambda(x-hz)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}}\right) dz + o\left(\frac{1}{n}\right).
\end{aligned}$$

Now set

$$u(z) = \frac{\lambda(x-z)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}} \tag{3.3}$$

and

$$z\lambda(x)^{\frac{1}{2}} = y \Leftrightarrow dz = \lambda(x)^{-\frac{1}{2}} dy. \tag{3.4}$$

Then we have

$$u\left(\frac{y}{\lambda(x)^{\frac{1}{2}}}\right) = \frac{\lambda\left(x-\frac{y}{\lambda^{\frac{1}{2}}(x)}\right)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}} \Leftrightarrow u\left(\frac{hy}{\lambda(x)^{\frac{1}{2}}}\right) = \frac{\lambda\left(x-\frac{hy}{\lambda(x)^{\frac{1}{2}}}\right)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}}. \tag{3.5}$$

Further set

$$\eta = \frac{h}{\lambda(x)^{1/2}} \tag{3.6}$$

and substitute back in (3.5) to get

$$\mathbb{E}\left\{\tilde{\lambda}_{n,2}(x|h)\right\} = \int \lambda(x)^{\frac{3}{2}} u^3(\eta y) K\Big(yu(\eta y)\Big) \lambda(x)^{-\frac{1}{2}}\, dy + o\left(\frac{1}{n}\right)$$
$$= \lambda(x) \int u^3(\eta y) K\Big(yu(\eta y)\Big)\, dy + o\left(\frac{1}{n}\right).$$

Thus the mean of $\tilde{\lambda}_{n,2}$ is of the form

$$\mathbb{E}\left\{\tilde{\lambda}_{n,2}(x|h)\right\} = \lambda(x) \int u^3(\eta y) K\Big(yu(\eta y)\Big)\, dy + o\left(\frac{1}{n}\right).$$

In order to analyze this mean expression further we expand $u(\eta y)$ in Taylor series around 0 and raise to the third power.

$$u(\eta y) = u(0) + \eta y u'(0) + \frac{(\eta y)^2}{2} u''(0) + \frac{(\eta y)^3}{3!} u'''(0) + \frac{(\eta y)^4}{4!} u^{(4)}(0) + \dots \tag{3.7}$$

Note that in this expansion we omitted the powers η^5 and above since they're not needed.

$$\begin{aligned} u^3(\eta y) \simeq\ & u^3(0) + \eta[3yu^2(0)u'(0)] + \eta^2 \left[3y^2 u(0)u'(0)^2 + \frac{3}{2} y^2 u^2(0) u''(0)\right] \\ & + \eta^3 \left[y^3 u'(0)^3 + \frac{3}{3!} y^3 u^2(0) u'''(0) + 3y^3 u'(0) u''(0)\right] \\ & + \eta^4 \left[y^4 u^2(0) u'(0) u'''(0) + \frac{3}{2} y^4 u'(0)^2 u''(0) + \frac{3}{4} y^4 u(0) u''(0)^2 + \frac{1}{8} y^4 u''''(0)\right] + \dots \end{aligned} \tag{3.8}$$

For the kernel, we multiply y with (3.7) and then expand the kernel with

$$x = y \ \text{ and } \ h = \eta y^2 u'(0) + \frac{\eta^2 y^3}{2} u''(0) + \dots$$

to get

$$K\Big(yu(\eta y)\Big) \simeq K(y) + \tag{3.9}$$

$$\left\{\eta y^2 u'(0) + \frac{\eta^2 y^3}{2} u''(0) + \dots\right\} K'(y) + \left\{\eta y^2 u'(0) + \frac{\eta^2 y^3}{2} u''(0) + \dots\right\}^2 \frac{K''(y)}{2} +$$
$$\left\{\eta y^2 u'(0) + \frac{\eta^2 y^3}{2} u''(0) + \dots\right\}^3 \frac{K'''(y)}{3!} + \left\{\eta y^2 u'(0) + \frac{\eta^2 y^3}{2} u''(0) + \dots\right\}^4 \frac{K''''(y)}{4!} \tag{3.10}$$

Rearranging (3.10) we get

$$K(y) + \eta\{y^2 u'(0)\} K'(y) + \eta^2 \left\{\frac{y^3}{2} u''(0) K'(y) + \frac{y^4 u'(0)^2}{2} K''(y)\right\} +$$
$$\eta^3 \left\{\frac{y^4}{3!} u'''(0) K'(y) + \frac{y^5}{2} u'(0) u''(0) K''(y) + \frac{y^6}{3!} u'(0)^3 K'''(y)\right\} +$$
$$\eta^4 \left\{\frac{y^5}{4!} u''''(0) K'(y) + \frac{y^6}{8} u''(0)^2 K''(y) + \frac{y^6}{6} u'(0) u^{(3)}(0) K''(y)\right\}^4 + \dots$$

Multiplying (3.8) with (3.10), noting that $u(0) = 1$ we get (A.6) in section A.2.1. We integrate (A.6) by integrating each one of the coefficients of the powers of the η's. Note that

$$\int y^2 K'(y) dy = 0$$

because we assume a symmetric kernel. Thus the coefficient of η is 0. Also the coefficient of η^3 is a sum of products of even and odd functions, thus its integral will be zero. For the coefficient of η^2 we calculate firstly the terms: $3y^2 u'(0)^2 K(y)$, $3y^3 u'(0)^2 K'(y)$, $\frac{y^4}{2} u'(0)^2 K''(y)$. We have

$$\int \frac{y^4}{2} K''(y)\, dy = \frac{y^4}{2} K'(y) - 2 \int y^3 K'(y)\, dy$$

thus,

$$\int 3y^2 K(y)\, dy + 3 \int y^3 K'(y)\, dy + \frac{y^4}{4} K'(y) - 2 \int y^3 K'(y)\, dy = \int 3y^2 K(y)\, dy + \int y^3 K'(y)\, dy + \frac{y^4}{4} K'(y).$$

Now,

$$\int 3y^2 K(y)\, dy + \left(y^3 K(y) - 3 \int y^2 K(y)\, dy \right) + \frac{y^4}{2} K'(y) = y^3 K(y) + \frac{y^4}{2} K'(y).$$

These three terms simplified give

$$\int 3y^2 K(y)\, dy + 3 \int y^3 K'(y)\, dy + \int \frac{y^4}{2} K''(y)\, dy = u'(0)^2 \Big[y^3 K(y) + \frac{y^4}{2} K'(y) \Big]. \quad (3.11)$$

Taking the remaining terms $\frac{3}{2} y^2 u''(0) K(y)$ and $\frac{y^3}{2} u''(0) K'(y)$ we have

$$\int \left\{ \frac{3}{2} y^2 K(y) + \frac{y^3}{2} K'(y) \right\} dy = \int \frac{3}{2} y^2 K(y)\, dy + \frac{y^3}{2} K(y) - \int \frac{3}{2} y^2 K(y)\, dy = \frac{y^3}{2} K(y). \quad (3.12)$$

From (3.11) and (3.12) the coefficient of η^2 is simplified to the following expression

$$u'(0)^2 \left[y^3 K(y) + \frac{y^4}{2} K'(y) \right] + u''(0) \left[\frac{y^3}{2} K(y) \right].$$

Since we assumed that the kernel K vanishes outside a compact interval, this quantity evaluated from $-\infty$ to ∞ will give zero. Now the coefficient of η^4 can be written as

$$\begin{aligned}
u'(0)^2 u''(0) &\left\{ \frac{3}{2} y^4 K(y) + \frac{9}{2} y^5 K'(y) + \frac{9}{4} y^6 K''(y) + \frac{1}{4} y^7 K'''(y) \right\} + \\
u''(0)^2 &\left\{ \frac{3}{4} y^4 K(y) + \frac{3}{4} y^5 K'(y) + \frac{1}{8} y^6 K''(y) \right\} + \\
u'(0)^4 &\left\{ y^5 K'(y) + \frac{3}{2} y^6 K''(y) + \frac{1}{2} y^7 K'''(y) + \frac{1}{4!} y^8 K''''(y) \right\} + \\
u'(0) u'''(0) &\left\{ y^4 K(y) + y^5 K'(y) + \frac{1}{6} y^6 K''(y) \right\} + u''''(0) \left\{ \frac{1}{8} y^4 K(y) + \frac{y^5}{4!} K'(y) \right\}.
\end{aligned}$$

By partial integration on each one of the square brackets we get the coefficient of η^4

$$\left(-\frac{1}{12}u''''(0) + u'(0)u'''(0) + 5u'^4(0) + \frac{3}{4}u''(0)^2 - 6u'(0)^2u''(0)\right)\int y^4K(y)\,dy.$$

Set the coefficient of the integral as $g_2(x)$. Calculating the derivatives of $u(z)$ and plugging them to the bias formula will give us the final expression for the bias. The derivatives of $u(z)$ at $z=0$ are

$$\begin{aligned}
u'(0) &= -\frac{\lambda'(x)}{2\lambda(x)},\\
u''(0) &= \frac{\lambda''(x)}{2\lambda(x)} - \frac{\lambda'(x)}{4\lambda^2(x)},\\
u'''(0) &= -\frac{\lambda'''(x)}{\lambda(x)} + \frac{3\lambda'(x)\lambda''(x)}{4\lambda^2(x)} - \frac{3\lambda'(x)^3}{\lambda^3(x)}\\
u''''(0) &= -\frac{15\lambda'(x)^4}{16\lambda^4(x)} + \frac{3\lambda'(x)\lambda''(x)}{\lambda^2(x)} - \frac{3\lambda''(x)^2}{\lambda^2(x)} - \frac{\lambda'(x)\lambda'''(x)}{\lambda^2(x)} + \frac{\lambda''''(x)}{2\lambda(x)}.
\end{aligned}$$

Substituting the values of the derivatives back to $g_2(x)$ gives

$$\begin{aligned}
g_2(x) = \frac{1}{12}&\left\{\frac{15\lambda'(x)^4}{16\lambda^4(x)} - \frac{3\lambda'(x)\lambda''(x)}{\lambda^2(x)} + \frac{3\lambda''(x)^2}{\lambda^2(x)} + \frac{\lambda'(x)\lambda'''(x)}{\lambda^2(x)} - \frac{\lambda''''(x)}{2\lambda(x)}\right\}\\
&+\left\{-\frac{\lambda'(x)}{2\lambda(x)}\right\}\left\{-\frac{\lambda'''(x)}{\lambda(x)} + \frac{3\lambda'(x)\lambda''(x)}{4\lambda^2(x)} - \frac{3\lambda'(x)^3}{\lambda^3(x)}\right\} + 5\left\{-\frac{\lambda'(x)}{2\lambda(x)}\right\}^4\\
&\qquad+\frac{3}{4}\left\{\frac{\lambda''(x)}{2\lambda(x)} - \frac{\lambda'(x)}{4\lambda^2(x)}\right\}^2 - 6\left\{-\frac{\lambda'(x)}{2\lambda(x)}\right\}^2\left\{\frac{\lambda''(x)}{2\lambda(x)} - \frac{\lambda'(x)}{4\lambda^2(x)}\right\}.
\end{aligned}$$

Simplifying, the above expression becomes

$$\begin{aligned}
g_2(x) = \frac{25}{64}\frac{\lambda'(x)^4}{\lambda^4(x)} &- \frac{3}{16}\frac{\lambda'(x)^2\lambda''(x)}{\lambda^3(x)} + \frac{1}{16}\frac{\lambda''(x)^2}{\lambda^2(x)} + \frac{1}{12}\frac{\lambda'(x)\lambda'''(x)}{\lambda^2(x)} - \frac{1}{24}\frac{\lambda''''(x)}{\lambda(x)}\\
&-\frac{1}{2}\frac{\lambda'(x)\left(-\frac{3}{8}\frac{\lambda'(x)^3}{\lambda^3(x)} + \frac{3}{4}\frac{\lambda'(x)\lambda''(x)}{\lambda^2(x)} - \frac{1}{2}\frac{\lambda'''(x)}{\lambda(x)}\right)}{\lambda(x)} + \frac{3}{4}\left(-\frac{1}{4}\frac{\lambda'(x)^2}{\lambda^2(x)} + \frac{1}{2}\frac{\lambda''(x)}{\lambda(x)}\right)^2\\
&\qquad-\frac{3}{2}\frac{\lambda'(x)^2\left(-\frac{1}{4}\frac{\lambda'(x)^2}{\lambda^2(x)} + \frac{1}{2}\frac{\lambda''(x)}{\lambda(x)}\right)}{\lambda^2(x)}.
\end{aligned}$$

Taking common factors and rearranging we find

$$\begin{aligned}
g_2(x) = &-\frac{-24\lambda'(x)^4 + 36\lambda'(x)^2\lambda''(x)^2\lambda(x) - 6\lambda''(x)^2\lambda^2(x)}{24\lambda^4(x)}\\
&\qquad+\frac{-8\lambda'(x)\lambda'''(x)\lambda^2(x) + \lambda''''(x)\lambda^3(x)}{24\lambda^4(x)}.
\end{aligned}$$

Recall that $\eta = h/\lambda^{\frac{1}{2}}$. Therefore the coefficient of η^4 is the coefficient of h^4/λ^2, i.e.

$$h^4 \frac{24\lambda'(x)^4 - 36\lambda'(x)^2\lambda''(x)^2\lambda(x) + 6\lambda''(x)^2\lambda^2(x) + 8\lambda'(x)\lambda'''(x)\lambda^2(x) - \lambda''''(x)\lambda^3(x)}{24\lambda^5(x)} \equiv g_1(x).$$

Hence

$$\mathbb{E}\left\{\tilde{\lambda}_{n,2}(x|h)\right\} = \lambda(x) + g_1(x)h^4 \int z^4 K(z)\, dz + o\left(h^4\right) + o\left(\frac{1}{n}\right).$$

As about the variance set,

$$\mathrm{I}_n(F) = \sum_{i=1}^{n} \frac{1}{n-i+1}\binom{n}{i-1} F(u)^{i-1}(1-F(u))^{n-i+1}.$$

and note that

$$n\mathrm{I}_n(F) \to \frac{1}{1-F} \text{ as } n \to +\infty$$

see for example Tanner and Wong [44]. Using this we find the leading term of the variance to be

$$\mathbb{E}\left\{\tilde{\lambda}_{n,2}(x|h)^2\right\} = \frac{1}{h^2}\int \sum_{i=1}^{n} \frac{\lambda(u)K^2\left(\frac{x-u}{h}\lambda^{\frac{1}{2}}(u)\right)}{n-i+1}\binom{n}{i-1} F(u)^{i-1}(1-F(u))^{n-i} f(u)\, du.$$

Notice that

$$\sum_{i=1}^{n} \frac{1}{n-i+1}\binom{n}{i-1} F(u)^{i-1}(1-F(u))^{n-i} = \frac{1}{1-F(u)}\mathrm{I}_n(F(u)).$$

Thus,

$$\mathbb{E}\left\{\tilde{\lambda}_{n,2}(x|h)^2\right\} = \frac{1}{nh^2}\int \lambda(u)^2 K^2\left(\frac{x-u}{h}\lambda(u)^{\frac{1}{2}}\right) n\mathrm{I}_n(F)\, du.$$

Set

$$Q(u) = \frac{\lambda(u)^2}{1-F(u)}.$$

Since the term $\{\mathbb{E}\tilde{\lambda}_{n,2}(x|h)\}^2$ is of much smaller order than $\mathbb{E}\{\tilde{\lambda}_{n,2}(x|h)^2\}$, asymptotically it is enough to consider only the $\mathbb{E}\{\tilde{\lambda}_{n,2}(x|h)^2\}$ of the variance. Thus,

$$\mathbb{V}\mathrm{ar}\left\{\tilde{\lambda}_{n,2}(x|h)\right\} \simeq \frac{1}{nh^2}\int Q(u)K^2\left(\frac{x-u}{h}\lambda(u)^{\frac{1}{2}}\right) du \text{ as } n \to +\infty.$$

Apply the change of variables $x - u = hz$. Then,

$$\begin{aligned}\frac{1}{nh^2}\int Q(u)K^2\left(\frac{x-u}{h}\lambda(u)^{\frac{1}{2}}\right) du &= \frac{1}{nh}\int Q(x-hz)K^2\left(z\lambda(x-hz)^{\frac{1}{2}}\right) dz \\ &= \frac{1}{nh}\int \frac{Q(x)}{Q(x)}Q(x-hz)K^2\left(z\frac{\lambda(x)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}}\lambda(x-hz)^{\frac{1}{2}}\right) dz.\end{aligned}$$

Now, set

$$\nu(z) = \frac{\lambda(x-z)^2}{1-F(x-z)} \left(\frac{\lambda(x)^2}{1-F(x)} \right)^{-1}.$$

Applying (3.4) and (3.6) yields

$$\nu\left(\frac{y}{\lambda(x)^{\frac{1}{2}}}\right) = \frac{\lambda\left(x - y\lambda(x)^{-\frac{1}{2}}\right)^2}{1-F\left(x - y\lambda(x)^{-\frac{1}{2}}\right)} \left(\frac{\lambda(x)^2}{1-F(x)}\right)^{-1} \Leftrightarrow$$

$$\nu\left(\frac{hy}{\lambda(x)^{\frac{1}{2}}}\right) = \frac{\lambda\left(x - hy\lambda(x)^{-\frac{1}{2}}\right)^2}{1-F\left(x - hy\lambda(x)^{-\frac{1}{2}}\right)} \left(\frac{\lambda(x)^2}{1-F(x)}\right)^{-1}$$

and so

$$\nu(\eta y) = \frac{\lambda(x-\eta y)^2}{1-F(x-\eta y)} \left(\frac{\lambda(x)^2}{1-F(x)} \right)^{-1}.$$

Using (3.3)-(3.6) we find

$$\begin{aligned}\int \frac{Q(x)}{Q(x)} Q(x-hz) K^2\left(z\frac{\lambda(x)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}} \lambda(x-hz)^{\frac{1}{2}} \right) dz &= \int Q(x)\lambda(x)^{-\frac{1}{2}} \nu(\eta y) K^2(yu(\eta y))\, dy \\ &= \frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)} \int \nu(\eta y) K^2(yu(\eta y))\, dy.\end{aligned}$$

The integral is approximated by Taylor expansions of u and K given by (3.7), (3.10) and expanding $\nu(\eta y)$ in Taylor series around 0 i.e.,

$$\nu(\eta y) = \nu(0) + \eta y \nu'(0) + \frac{(\eta y)^2}{2}\nu''(0) + \frac{(\eta y)^3}{3!}\nu'''(0) + \frac{(\eta y)^4}{4!}\nu^{(4)}(0) + \dots \ .$$

Squaring (3.10) and substituting to the integral, we find that the asymptotic variance is

$$\mathbb{V}\mathrm{ar}\left\{\tilde{\lambda}_{n,2}(x|h)\right\} = \frac{1}{nh}\frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)} \int K^2(t)\, dt + o\left(\frac{1}{nh}\right). \quad \blacksquare$$

Remark 3.1. In the above proof, the bias of $\tilde{\lambda}_{n,2}$ is obtained by direct calculation. However we point out that one may apply the theorem in Hall [14] to derive the same bias expression. In particular Hall's theorem applies to an estimator of the form

$$\hat{\theta}(x) = \frac{1}{nh}\sum_{i=1}^{n} Y_i a(X_i) K^{(t)}\left(\frac{x-X_i}{h} a(X_i)\right)$$

where (X_i, Y_i) is a data pair, $K^{(t)}$ denotes the tth derivative of the kernel and $h_i = h/a(X_i)$ is the bandwidth associated with (X_i, Y_i). By setting

$$Y_i = \frac{1}{1-F_n(X_i)}, \quad t = 0 \ \text{ and } \ a(X_i) = \lambda^{1/2}(X_i)$$

we get $\hat{\theta}(x) = \tilde{\lambda}_{n,2}(x)$.

Remark 3.2. From the above theorem the mean square error distance between $\tilde{\lambda}_{n,2}(x|h)$ and $\lambda(x)$ can easily shown to be $O(n^{-\frac{8}{9}})$. For this first note that the optimal value of h (in the bandwidth function, $h\lambda^{\frac{1}{2}}(x)$) which minimizes mean square error of $\tilde{\lambda}_{n,2}(x|h)$ is asymptotically constant multiple of $n^{-\frac{1}{9}}$. Then, for such h,

$$\mathbb{E}\left\{\tilde{\lambda}_{n,2}(x|h) - \lambda(x)\right\}^2 = O(n^{-\frac{8}{9}}).$$

Also with $h \sim n^{-\frac{1}{9}}$ one could establish that as $n \to +\infty$

$$\frac{\sqrt{nh}\left(\tilde{\lambda}_{n,2}(x|h) - \mathbb{E}\tilde{\lambda}_{n,2}(x|h)\right)}{\sigma(x)} \to N(0,1)$$

where

$$\sigma^2(x) = \frac{\lambda(x)^{\frac{3}{2}}}{1 - F(x)} \int K^2(t)\, dt.$$

thus giving us the distance between $\tilde{\lambda}_{n,2}(x|h)$ and $\lambda(x)$ to be $O_p(n^{-\frac{4}{9}})$. For an outline see the proof of lemma 3.6.1. Similar results are true for $\tilde{\lambda}_{n,1}$ as well.

Remark 3.3. In the discussion above, the bandwidths of the ideal estimators (i.e. h in (3.1) and (3.2)) are denoted and treated as constants. The case of random bandwidths $\hat{h}$ in (3.1) and (3.2) could be analyzed as follows. Assume that the kernel is compactly supported and twice continuously differentiable and that $\hat{h}$ satisfies $n^{\frac{1}{9}}\hat{h} \to c$ in probability, where $0 < c < +\infty$. Then, we have

$$\begin{aligned} \tilde{\lambda}_{n,1}(x|\hat{h}) &= \tilde{\lambda}_{n,1}(x|cn^{-\frac{1}{9}}) + o_p(n^{-\frac{4}{9}}) \\ \text{and} \;\; \tilde{\lambda}_{n,2}(x|\hat{h}) &= \tilde{\lambda}_{n,2}(x|cn^{-\frac{1}{9}}) + o_p(n^{-\frac{4}{9}}) \end{aligned}$$

as $n \to +\infty$ for every $x \in (0, T)$. Recall estimator $\bar{\lambda}$ defined on page 36. Now, since

$$\left|\tilde{f}(x|\hat{h}) - \tilde{f}(x|cn^{-\frac{1}{9}})\right| = o_p\left(n^{-\frac{4}{9}}\right)$$

(see Hall and Marron [16]), then the proof of $\tilde{\lambda}_{n,1}$ is done simply by using the strong convergence of F_n to F to replace $\tilde{\lambda}_{n,1}(x|\hat{h})$ and $\tilde{\lambda}_{n,1}(x|cn^{-\frac{1}{9}})$ by $\bar{\lambda}(x|\hat{h})$ and $\bar{\lambda}(x|cn^{-\frac{1}{9}})$ respectively. In the case of $\tilde{\lambda}_{n,2}$ we can use an argument similar to that of Abramson [1] for the proof of equation (2) there. Adopting Abramson's notation, the core of the argument is to regard $\hat{\lambda}_{n,2}$ as a continuous path stochastic process and to form the sequence, say Y_n, which in our case is the difference between $\hat{\lambda}_{n,2}$ and $\lambda(x)$ inflated by $n^{\frac{2}{5}}$. Since tightness of Y_n implies its equicontinuity and since Y_n is bounded we can use the Arzela-Ascoli theorem and the dominated convergence theorem to find a bound for the mean of the sequence $Z_n^a = Y_n^2 \wedge a$ where a is a fixed positive number (see [1], section 2). Tightness of Y_n will follow by using the same modification of Billingsley's theorem (see [1], section 3), with calculations analogous to those used in the proof of theorem 3.3.2.

3.4 Adaptive estimators.

Recall that $\tilde{\lambda}_{n,1}$ and $\tilde{\lambda}_{n,2}$ are 'ideal' estimators in the sense that the probability density function f and hazard rate function λ involved in the bandwidth functions for $\tilde{\lambda}_{n,1}$ and $\tilde{\lambda}_{n,2}$ respectively, are the true and thus unknown functions. In order to get practically useful estimators we replace f and λ by their simple kernel estimators. This leads to the definition of the so-called *adaptive* estimators for the hazard rate function. In the case of $\tilde{\lambda}_{n,1}$ its adaptive version is defined as

$$\hat{\lambda}_{n,1}(x|h_1,h_2) = \frac{\hat{f}(x|h_1,h_2)}{1-F_n(x)} \tag{3.13}$$

where

$$\hat{f}(x|h_1,h_2) = \frac{1}{nh_2}\sum_{i=1}^{n}\hat{f}(X_i|h_1)^{\frac{1}{2}}K\left(\frac{x-X_i}{h_2}\hat{f}(X_i|h_1)^{\frac{1}{2}}\right)$$

and $\hat{f}(\cdot|h)$ is as in chapter 1. The adaptive counterpart of $\tilde{\lambda}_{n,2}$ is defined as

$$\hat{\lambda}_{n,2}(x|h_1,h_2) = \frac{1}{h_2}\sum_{i=1}^{n}\frac{\hat{\lambda}(X_{(i)}|h_1)^{\frac{1}{2}}K\left(\frac{x-X_{(i)}}{h_2}\hat{\lambda}(X_{(i)}|h_1)^{\frac{1}{2}}\right)}{n-i+1} \tag{3.14}$$

where $\hat{\lambda}$ is the 'pilot' estimator. Here, as pilot we use estimator $\hat{\lambda}_2$ defined in (1.3).

Although there is no theoretical reason why the kernels of the pilot and the adaptive estimators should be the same, here we use same kernels for the sake of simplicity. Note that for $\hat{\lambda}_{n,j}$, $j=1,2$ based on random bandwidths we replace h_1,h_2 in the definition of $\hat{\lambda}_{n,j}$, $j=1,2$ by $\hat{h}_1$ and $\hat{h}_2$. Next we concentrate on quantifying the distance between the ideal and adaptive estimators.

3.5 Comparison of the ideal and adaptive estimators.

3.5.1 Comparison of the ordered estimators.

The main objective of this subsection is to prove that up to terms of $o_p(n^{-4/9})$ the adaptive equals the ideal plus a remainder term. In the first of the next two theorems we show that the remainder term is a random variable, and that it is asymptotically normally distributed is established in the second theorem.

Theorem 3.5.1. *Assume that the kernel satisfies conditions A1, A2 and that it has two bounded derivatives. Suppose that $\lambda > 0$ is three times differentiable with the third derivatives satisfying a Lipschitz condition of unit order. Also, we assume that if the bandwidth is random then, with probability 1 as $n \to +\infty$*

$$n^{-a-\frac{1}{5}} < \hat{h}_1 < n^{a-\frac{1}{5}}, \text{ where } a < \frac{1}{5}$$
$$\text{and } \eta n^{-\frac{1}{9}} < \hat{h}_2 < \rho n^{\frac{1}{9}}$$

with $\rho > \eta > 0$. Then

$$\hat{\lambda}_{n,2}(x|\hat{h}_1, \hat{h}_2) = \tilde{\lambda}_{n,2}(x|\hat{h}_2) + \mathrm{T}_2(x|\hat{h}_1, \hat{h}_2) + o_p\left(n^{-\frac{4}{9}}\right) \tag{3.15}$$

where

$$\mathrm{T}_2(x|h_1, h_2) = \frac{1}{2nh_1h_2} \sum_{i=1}^{n} \frac{t_2(X_i, x|h_1, h_2)}{1 - F(X_i)},$$

$$t_2(u, x|h_1, h_2) = \mathbb{E}\left\{\lambda(X_i)^{-\frac{1}{2}} \left\{\frac{K\left(\frac{X_i - u}{h_1}\right)}{1 - F(X_i)} - h_1\mu_2(X_i|h_1)\right\} \mathrm{L}\left\{\left(\frac{x - X_i}{h_2}\right) \lambda(X_i)^{\frac{1}{2}}\right\}\right\},$$

and μ_2 is defined to be $\mu_2(x|h) = \mathbb{E}\hat{\lambda}_2(x|h)$.

Define

$$\mathrm{S}_2(x|h_2) = \tilde{\lambda}_{n,2}(x|h_2) - \mathbb{E}\tilde{\lambda}_{n,2}(x|h_2)$$

and

$$\mathrm{L}_1(z) = zK'(z), \quad \mathrm{L} = K + \mathrm{L}_1.$$

The next theorem shows that the remainder term, T_2, is of the same order as the difference between $\tilde{\lambda}_{n,2}$ and the true hazard λ.

Theorem 3.5.2. *Assume that the kernel satisfies conditions A1, A2 and that it has two bounded derivatives. Let $\lambda > 0$ be bounded on $(0, T)$ and continuous at $x \in (0, T)$. Suppose that for non-random bandwidths h_1, h_2 satisfying $n^{\varepsilon} \max(h_1, h_2) \to 0$, $n^{1-\varepsilon} \min(h_1, h_2) \to \infty$ for some $\varepsilon > 0$ and $h_1 h_2^{-1} \to 0$ we have $\hat{h}_1/h_1 \to 1$ and $\hat{h}_2/h_2 \to 1$ in probability. Then*

$$\sqrt{nh_2}\left(\mathrm{S}_2(x|\hat{h}_2), \mathrm{T}_2(x|\hat{h}_1, \hat{h}_2)\right) \to (N_1, N_2)$$

where (N_1, N_2) is a bivariate normal distribution with mean 0 and covariance

$$\mathbb{V}\mathrm{ar}(N_1) = \frac{\lambda(x)^{\frac{3}{2}}}{1 - F(x)} \int K^2,$$

$$\mathbb{V}\mathrm{ar}(N_2) = \frac{1}{4} \frac{\lambda(x)^{\frac{3}{2}}}{1 - F(x)} \int \mathrm{L}^2,$$

$$\mathbb{C}\mathrm{ov}(N_1, N_2) = \frac{1}{2} \frac{\lambda(x)^{\frac{3}{2}}}{1 - F(x)} \int K\mathrm{L}.$$

An immediate conclusion that can be drawn is that the remainder term does not lessen the rate of convergence, however, it does prevent the adaptive estimator from achieving the same first-order properties as the ideal. Also, note the variance formulae of N_1 and N_2. The presence of $1 - F(x)$ in the denominator means that the variance increases with x. However, as both the kernel and the hazard rate are bounded, the distance between the ideal and the adaptive estimators will be at an acceptable level.

Remark 3.4. Since, for large n, $\mathbb{E}N_1 = \mathbb{E}N_2 = 0$ we conclude that $\hat{\lambda}_{n,2}$ and $\tilde{\lambda}_{n,2}$ have the same asymptotic bias. Now, notice that

$$\int K(z)\mathrm{L}(z)\, dz = \int \left(K^2(z) + zK(z)K'(z)\right)\, dz$$

and

$$\int zK(z)K'(z)\,dz = zK^2(z) - \int \left(K^2(z) + zK(z)K'(z)\right)\,dz$$

and therefore

$$2\int zK(z)K'(z)\,dz = zK^2(z) - \int K^2(z)\,dz.$$

Since we assume that the kernel vanishes outside a compact set,

$$\int K(z)\mathrm{L}(z)\,dz = \frac{1}{2}\int K^2(z)\,dz > 0.$$

Thus the asymptotic covariance of S_2 and T_2 is assuredly positive. Hence, the presence of the remainder term T_2 in the RHS of (3.15) we conclude that the adaptive estimator has larger asymptotic variance than the ideal.

Remark 3.5. As can be seen from theorem 3.5.2 the term $\mathrm{T}_2(x|\hat{h}_1, \hat{h}_2)$ asymptotically does not depend on $\hat{h}_1$. Therefore, the use as pilot of a more accurate estimator than $\hat{\lambda}_1$ or $\hat{\lambda}_2$ will not improve the asymptotic properties of $\hat{\lambda}_{n,2}$ and $\tilde{\lambda}_{n,2}$. Moreover in situations where computational speed is an issue, the use of a more precise pilot estimator is not advisable since the improvement gained won't be worthy the computational cost added.

Remark 3.6. The condition $h_1 h_2^{-1} \to 0$ of theorem 3.5.2 expresses the way we should choose the bandwidths. Practically we would choose a bigger value for h_2 than for h_1, i.e. $h_1 \sim n^{-\frac{1}{5}}$ and $h_2 \sim n^{-\frac{1}{9}}$ because in this way we achieve the fastest rate of convergence to the true hazard rate. In this sense the condition is essential for the validity of the theorem.

Remark 3.7. Selection of sets

$$\mathcal{H}_1 = (n^{-a-\frac{1}{5}}, n^{a-\frac{1}{5}}) \;\text{ and }\; \mathcal{H}_2 = (\eta n^{-\frac{1}{9}}, \rho n^{\frac{1}{9}})$$

for the bandwidths is justified by the standard theory in data driven bandwidths ([26], [18]) and the theoretical work of Jones in [19], which suggest that we can achieve optimal rate of convergence if we choose $\hat{h}_1, \hat{h}_2$ so that

$$\mathrm{P}\left(n^{-a} < n^{\frac{1}{5}}\hat{h}_1 < n^a\right) \to 1,$$
$$\lim_{\eta\to 0,\rho\to\infty}\ \lim_{n\to\infty}\inf \mathrm{P}\left(\eta < n^{\frac{1}{9}}\hat{h}_2 < \rho\right) = 1.$$

Such data driven bandwidths can be found by following for example the methods of Park and Marron in [26]. However note that a key assumption is that we choose h_1 to be close to optimal, otherwise (3.15) will fail.

3.5.2 Comparison of the ratio estimators.

Using relationship (1.5) between $\hat{f}(x|h_1,h_2)$ and $\tilde{f}(x|h_2)$ in Hall and Marron [16] we set out to show that a similar relationship holds for $\tilde{\lambda}_{n,1}$ and $\hat{\lambda}_{n,1}$. This is expressed in the form of the following two theorems. First, define

$$S_1(x|h_2) = \bar{\lambda}_{n,1}(x|h_2) - \mathbb{E}\bar{\lambda}_{n,1}(x|h_2).$$

and

$$T_1(x|h_1,h_2) = \frac{1}{2nh_1h_2}\sum_{i=1}^{n}\frac{t_1(X_i,x|h_1,h_2)}{1-F(x)},$$

where

$$t_1(u,x|h_1,h_2) = \mathbb{E}\left\{f(X_i)^{-\frac{1}{2}}\left\{K\left(\frac{X_i-u}{h_1}\right) - h_1\mu_1(X_i|h_1)\right\}\mathrm{L}\left\{\left(\frac{x-X_i}{h_2}\right)f(X_i)^{\frac{1}{2}}\right\}\right\},$$

and μ_1 is defined to be $\mu_1(x|h) = \mathbb{E}\tilde{f}(x|h)$.

The first theorem shows that up to terms of $o_p(n^{-4/9})$ the adaptive equals the ideal plus a remainder term.

Theorem 3.5.3. *Assume that the kernel satisfies conditions A1, A2 and in addition that it has two bounded derivatives. Suppose and that the function f has three derivatives of all types, that the third derivatives satisfy a Lipschitz condition of unit order, and that f is bounded away from zero on $[0,T]$. Then*

$$\hat{\lambda}_{n,1}(x|\hat{h}_1,\hat{h}_2) = \tilde{\lambda}_{n,1}(x|\hat{h}_2) + T_1(x|\hat{h}_1,\hat{h}_2) + o_p\left(n^{-\frac{4}{9}}\right)$$

where the data-driven bandwidths $\hat{h}_1$ and $\hat{h}_2$ take values in the sets

$$\mathcal{H}_1 \equiv \{\hat{h}_1 : n^{-a-\frac{1}{5}} \le \hat{h}_1 \le n^{a-\frac{1}{5}}\},$$
$$\mathcal{H}_2 \equiv \{\hat{h}_2 : \eta n^{-\frac{1}{9}} \le \hat{h}_2 \le \rho n^{-\frac{1}{9}}\}$$

with $\eta > \rho > 0$.

From practical point of view this theorem shows that the adaptive estimator is almost but not quite as good as, the ideal estimator. The following theorem shows that the distance, $T_1(x|\hat{h}_1,\hat{h}_2)$, between the two estimators is a random variable asymptotically normally distributed.

Theorem 3.5.4. *Assume that the kernel satisfies conditions A1,A2 and in addition that it has two bounded derivatives. Suppose that f is bounded on $[0,T]$ and continuous at $x \in [0,T]$. Suppose also that for non-random bandwidths h_1 and h_2 satisfying $n^{\varepsilon}\max(h_1,h_2) \to 0$, $n^{1-\varepsilon}\min(h_1,h_2) \to \infty$ for some $\varepsilon > 0$ and $h_1h_2^{-1} \to 0$ we have $\hat{h}_1/h_1 \to 1$ and $\hat{h}_2/h_2 \to 1$ in probability. Then,*

$$\sqrt{nh_2}\left(S_1(x|\hat{h}_2), T_1(x|\hat{h}_1,\hat{h}_2)\right) \to (N_1,N_2)$$

where (N_1, N_2) is a bivariate normal distribution with mean 0 and variance

$$\begin{aligned}\mathbb{V}\mathrm{ar}(N_1) &= \frac{f(x)^{3/2}}{(1-F(x))^2}\int K^2,\\ \mathbb{V}\mathrm{ar}(N_2) &= \frac{1}{4}\frac{f(x)^{3/2}}{(1-F(x))^2}\int \mathrm{L}^2,\\ \mathbb{C}\mathrm{ov}(N_1,N_2) &= \frac{1}{2}\frac{f(x)^{3/2}}{(1-F(x))^2}\int K\mathrm{L}.\end{aligned}$$

We note here that remarks analogous to those of subsection 3.5.1 can be made for these two theorems as well. The proofs of both theorems follow from theorems 3.1 and 3.2 of Hall and Marron [16] and therefore we omit the details.

3.6 Proof of theorems 3.5.1 and 3.5.2.

3.6.1 Proof of theorem 3.5.1.

Observe that

$$\begin{aligned}\hat{\lambda}_{n,2}(x|\hat{h}_1,\hat{h}_2) &= \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}K\left(\frac{x-X_i}{h}\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}\right)}{1-F(X_i)}\\ &\quad+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}K\left(\frac{x-X_i}{h}\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}\right)\left\{\frac{1}{1-F_n(X_i)}-\frac{1}{1-F(X_i)}\right\}\\ &= \bar{\lambda}_{n,2}(x|\hat{h}_1,\hat{h}_2)+o_p(n^{-\frac{1}{2}})\end{aligned}$$

where

$$\bar{\lambda}_{n,2}(x|\hat{h}_1,\hat{h}_2) = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}K\left(\frac{x-X_i}{\hat{h}_2}\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}\right)}{1-F(X_i)}.$$

Also,

$$\begin{aligned}\tilde{\lambda}_{n,2}(x|\hat{h}_2) &= \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}K\left(\frac{x-X_i}{h}\lambda(X_i)^{\frac{1}{2}}\right)}{1-F(X_i)}\\ &\quad+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\lambda(X_i)^{\frac{1}{2}}K\left(\frac{x-X_i}{h}\lambda(X_i)^{\frac{1}{2}}\right)\left\{\frac{1}{1-F_n(X_i)}-\frac{1}{1-F(X_i)}\right\}.\end{aligned}$$

Hence, we can write

$$\tilde{\lambda}_{n,2}(x|\hat{h}_2) = \bar{\lambda}_{n,2}(x|\hat{h}_2)+o_p(n^{-\frac{1}{2}})$$

with

$$\bar{\lambda}_{n,2}(x|\hat{h}_2) = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}K\left(\frac{x-X_i}{\hat{h}_2}\lambda(X_i)^{\frac{1}{2}}\right)}{1-F(X_i)}.$$

Thus, (3.15) is equivalent to

$$\bar{\lambda}_{n,2}(x|\hat{h}_1, \hat{h}_2) = \bar{\lambda}_{n,2}(x|\hat{h}_2) + \mathrm{T}_2(x|\hat{h}_1, \hat{h}_2) + o_p\left(n^{-\frac{4}{9}}\right). \tag{3.16}$$

Now, (3.16) is implied by

$$\sup_{x\in[0,T]} \left|\bar{\lambda}_{n,2}(x|\hat{h}_1, \hat{h}_2) - \bar{\lambda}_{n,2}(x|\hat{h}_2) - \mathrm{T}_2(x|\hat{h}_1, \hat{h}_2)\right| = o_p\left(n^{-\frac{4}{9}}\right). \tag{3.17}$$

Applying the definition of stochastic convergence we can rewrite (3.17) as

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{x\in[0,T]} \left|\bar{\lambda}_{n,2}(x|\hat{h}_1, \hat{h}_2) - \bar{\lambda}_{n,2}(x|\hat{h}_2) - \mathrm{T}_2(x|\hat{h}_1, \hat{h}_2)\right| > \xi n^{-\frac{4}{9}}\right) = 0.$$

To prove this, we define a relationship between the pilot estimate and the true failure rate, and we use this to write the adaptive estimate in terms of the ideal and some remainder terms. Then we work out the difference between these terms and T_2. First we split the adaptive into parts and then we investigate how T_2 is related to those remaining terms. Lemmas used throughout the proof are proved in section 3.7. Define $\delta(x)$ from the relation

$$\hat{\lambda}_2(x|\hat{h}_1)^{\frac{1}{2}} = \lambda(x)^{\frac{1}{2}}\{1 + \delta(x)\}. \tag{3.18}$$

Also set

$$b(x|h) = \mu_2(x|h) - \lambda(x), \text{and } \mathrm{D}(x|h) = \hat{\lambda}_2(x|h) - \mu_2(x|h).$$

Then we can write (lemma 3.7.1, pp. 62)

$$\hat{\lambda}_2(x|\hat{h}_1)^{\frac{1}{2}} = \lambda(x)^{\frac{1}{2}} \left\{1 + \frac{\mathrm{D}(x|\hat{h}_1) + b(x|\hat{h}_1)}{\lambda(x)}\right\}^{\frac{1}{2}}. \tag{3.19}$$

Next we break the adaptive into parts by performing a Taylor expansion on the kernel, substituting back to the estimator and rearrange. Using (3.18) and Taylor series we write the kernel of the adaptive as

$$\begin{aligned}
K\left\{\left(\frac{x-u}{\hat{h}_2}\right)\hat{\lambda}_2(u|\hat{h}_1)^{\frac{1}{2}}\right\} &= K\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}(1+\delta(u))\right\} \\
&= K\left\{\left(\frac{x-u}{\hat{h}_2}\right)\left(\lambda(u)^{\frac{1}{2}} + \lambda(u)^{\frac{1}{2}}\delta(u)\right)\right\} \\
&= K\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\right\} \\
&\quad + \left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\delta(u)K'\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\right\} + \delta_2(x,u) \\
&= K\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\right\} + \delta(u)\mathrm{L}_1\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\right\} \\
&\quad + \delta_2(x,u)
\end{aligned} \tag{3.20}$$

where

$$|\delta_2(x,u)| \le C_1\delta(u)^2 \mathrm{I}\left(|x-X_i| \le C_2\hat{h}_2\right)$$

uniformly in $x, u \in (0,T)$. Here the indicator function $\mathrm{I}\left(|x-X_i| \le C_2\hat{h}_2\right)$ is introduced in order to exclude large values of u as this will ensure that observations away from the evaluation point will not have large effect on the estimate. Substituting this expression for the kernel to $\bar{\lambda}_{n,2}$ gives

$$\bar{\lambda}_{n,2}(x|\hat{h}_1,\hat{h}_2) =$$

$$\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\hat{\lambda}_2(X_i|\hat{h}_1)^{\frac{1}{2}}\left(K\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}+\delta(X_i)\mathrm{L}_1\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}+\delta_2(x,X_i)\right)}{1-F(X_i)}$$

$$=\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\hat{\lambda}_2(X_i|\hat{h}_1)^{\frac{1}{2}}K\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}+$$

$$\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\hat{\lambda}_2(X_i|\hat{h}_1)^{\frac{1}{2}}\delta(X_i)\mathrm{L}_1\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\hat{\lambda}_2(X_i|\hat{h}_1)^{\frac{1}{2}}}{1-F(X_i)}\delta_2(x,X_i).$$

Now, using (3.18)

$$\begin{aligned}\bar{\lambda}_{n,2}(x|\hat{h}_1,\hat{h}_2) &= \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\left(\lambda(X_i)^{\frac{1}{2}}+\delta(X_i)\lambda(X_i)^{\frac{1}{2}}\right)K\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\ &+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\left(\lambda(X_i)^{\frac{1}{2}}+\delta(X_i)\lambda(X_i)^{\frac{1}{2}}\right)\delta(X_i)\mathrm{L}_1\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\ &+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{(1+\delta(X_i))\,\lambda(X_i)^{\frac{1}{2}}\delta_2(x,X_i)}{1-F(X_i)}.\end{aligned}$$

Rearranging,

$$\begin{aligned}\bar{\lambda}_{n,2}(x|\hat{h}_1,\hat{h}_2) &= \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}K\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\ &+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}(\delta(X_i)+\delta^2(X_i))\mathrm{L}_1\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\ &+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\delta(X_i)\lambda(X_i)^{\frac{1}{2}}K\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\ &+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}+\delta(X_i)\lambda(X_i)^{\frac{1}{2}}}{1-F(X_i)}\delta_2(x,X_i).\end{aligned}$$

Hence,

$$\bar{\lambda}_{n,2}(x|\hat{h}_1,\hat{h}_2)=\bar{\lambda}_{n,2}(x|\hat{h}_2)+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}\delta(X_i)\mathrm{L}\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}+\delta_3(x) \tag{3.21}$$

where

$$\delta_3(x) = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\left\{\frac{\delta^2(X_i)\mathrm{L}_1\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)} + \frac{\delta(X_i)\lambda(X_i)^{\frac{1}{2}}K\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\right.$$
$$\left.+\frac{\lambda(X_i)^{\frac{1}{2}}+\delta(X_i)\lambda(X_i)^{\frac{1}{2}}}{1-F(X_i)}\delta_2(x,X_i)\right\} \le C_1\left\{\sup_{X_i\in[0,\mathrm{T}]}\delta^2(X_i)\right\}\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\mathrm{I}\left(|x-X_i|\le C_2\hat{h}_2\right)$$

for a suitably chosen constant C_1 which includes the bound for the denominator, uniformly in $x \in (0,T)$. Since we already have $\bar{\lambda}_{n,2}$ in (3.21), we only need to form the T_2 term from the last two terms of (3.21). It will be easier if we use functions similar to those used in already existing theorems (c.f. [16]).

Rewrite (3.19) as

$$\hat{\lambda}_2(x|\hat{h}_1)^{\frac{1}{2}} = \lambda(x)^{\frac{1}{2}}\left\{1+\delta_4(x)+\delta_5(x)\right\}$$

where

$$\delta_4(x) = \frac{\mathrm{D}(x|\hat{h}_1)+b(x|\hat{h}_1)}{2\lambda(x)}, \quad \delta_5(x) \le C\left\{\mathrm{D}(x|\hat{h}_1)^2 + b(x|\hat{h}_1)^2\right\}$$

uniformly in $x \in (0,T)$. Then, for $\delta = \delta_4+\delta_5$ (lemma 3.7.2, pp. 62)

$$\bar{\lambda}_{n,2}(x|\hat{h}_1,\hat{h}_2) = \bar{\lambda}_{n,2}(x|\hat{h}_2) + \frac{1}{2}\varepsilon_1(x|\hat{h}_1,\hat{h}_2) + \frac{1}{2}\varepsilon_2(x|\hat{h}_1,\hat{h}_2) + \varepsilon_3(x|\hat{h}_1,\hat{h}_2) \tag{3.22}$$

where we define the ε_i, $i=1,2,3$ to be

$$\begin{aligned}
\varepsilon_1(x|\hat{h}_1,\hat{h}_2) &= \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{-\frac{1}{2}}\mathrm{D}(X_i|\hat{h}_1)\mathrm{L}\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\
\varepsilon_2(x|\hat{h}_1,\hat{h}_2) &= \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{-\frac{1}{2}}b(X_i|\hat{h}_1)\mathrm{L}\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\
|\varepsilon_3(x|\hat{h}_1,\hat{h}_2)| &\le C_1\varepsilon_4(x|\hat{h}_1,\hat{h}_2)\\
&\equiv \frac{C_1}{n\hat{h}_2}\left(\sup_{x\in[0,T]}\left\{\mathrm{D}(x|\hat{h}_1)^2 + b(x|\hat{h}_1)^2\right\}\sum_{i=1}^{n}\mathrm{I}\left(|x-X_i|\le C_2\hat{h}_2\right)\right).
\end{aligned}$$

Since we have proved so far that the adaptive estimator can be written in the form (3.22) we now need to work out the sum of the ε_i's, $i=1,2,3$. Specifically we want to show that

$$\frac{\sup_{x\in[0,T]}\left|\frac{1}{2}\varepsilon_1(x|\hat{h}_1,\hat{h}_2)+\frac{1}{2}\varepsilon_2(x|\hat{h}_1,\hat{h}_2)+\varepsilon_3(x|\hat{h}_1,\hat{h}_2)-\mathrm{T}_2(x|\hat{h}_1,\hat{h}_2)\right|}{n^{-\frac{4}{9}}} \xrightarrow{p} 0. \tag{3.23}$$

Equivalently (lemma 3.7.3, pp. 62), we show that

$$\sup_{x\in[0,T],h_1\in\mathcal{H}_1,h_2\in\mathcal{H}_2}\mathrm{P}\left\{|\varepsilon_i(x|h_1,h_2)| > \xi n^{-\frac{4}{9}}\right\} = O(n^{-r}), \quad r>0 \tag{3.24}$$

for $i=2,4$ and

$$\sup_{x\in[0,T],h_1\in\mathcal{H}_1,h_2\in\mathcal{H}_2} \mathrm{P}\left\{|\varepsilon_1(x|h_1,h_2) - 2\mathrm{T}_2(x|h_1,h_2)| > \xi n^{-\frac{4}{9}}\right\} = O(n^{-r}). \tag{3.25}$$

A crucial point of the proof will be the repeated use of inequality (21.5) of Burkholder [6]. In Burkholder's notation let $(f_1, f_2, \dots)$ be a martingale relative to $(\mathcal{F}_1, \mathcal{F}_2, \dots)$, a nondecreasing sequence of sub-σ-fields of $\mathcal{F}$. Let $(d_1, d_2, \dots)$ be the difference sequence of $f : f_n = \sum_{i=1}^{n} d_i$, $n \geq 1$. The square function and the maximal function of f are

$$S(f) = \left(\sum_{i=1}^{+\infty} \mathbb{E}\left\{d_i^2\right\}\right)^{\frac{1}{2}} \quad \text{and} \quad f^* = \sup_n |f_n|$$

respectively. Then, for a function Φ that is non-decreasing and continuous on $[0, +\infty)$ and satisfies $\Phi(2x) \leq c\Phi(x)$ the following inequality holds

$$\mathbb{E}\Phi(f^*) \leq c\mathbb{E}\Phi(S(f)) + C\sum_{i=1}^{+\infty} \mathbb{E}\Phi\left(|d_i|\right) \tag{3.26}$$

for some constant C. Starting with the case $i = 4$ we have

case i=4: Looking at the structure of ε_4 we see that we need to bound functions D, b and the sum. We start with function D. Suppose that $\mathcal{J}$ is a finite set, subset of $(0, \mathrm{T})$ such that

$$\mathcal{J} = \{x \in (0,T) : \text{for some } y \in [0,T], |x-y| < \varepsilon\}, \varepsilon > 0$$

An upper bound for $\mathrm{E}|\mathrm{D}(x|h_1)|^l$ for some appropriate constant C_1 will be (lemma 3.7.4, pp. 63)

$$\mathbb{E}\left\{|\mathrm{D}(x|h_1)|^l\right\} \leq C_1\left(\frac{1}{nh_1}\right)^{\frac{l}{2}} \leq C_1\left(\frac{1}{n^{a+\frac{4}{5}}}\right)^{\frac{l}{2}}$$

uniformly in $x \in (0,T)$.

Suppose that the number of the elements of $\mathcal{J}$ increases at most algebraically fast in n. That is we assume that the set $\mathcal{J}$ has $O(n^s)$ elements, where s is large but fixed. Working as in Stone [43] and using the fact that $\sum_i \mathbb{E}z_i \leq \{\#\mathcal{J}\} \sup \mathbb{E}z_i$ together with corollary 2.2 from Hall and Heyde ([15], pp. 19) we prove that the probability below is $O(n^{-r})$. Then we use Hölder continuity of the kernel to extend the result in the general case. We have

$$\mathrm{P}\left\{\sup_{x\in\mathcal{J}} \mathrm{D}(x|h_1)^2 > \xi n^{-\frac{4}{9}}\right\} \leq \{\#\mathcal{J}\}\left\{\xi^{-1}n^{\frac{4}{9}}\right\}^{\frac{1}{2}} \sup_{x\in\mathcal{J}} \mathbb{E}\left\{|\mathrm{D}(x|h_1)|^l\right\}.$$

Since the kernel is a Hölder continuous function we have that

$$|\mathrm{D}(x|h_1) - \mathrm{D}(y|h_1)| \leq c\,|x-y|^l.$$

From the definition of Hölder continuity we have that l can be any number greater than zero. This means that the distance between two successive elements of $\mathcal{J}$ can be as small or as big we like. Therefore, on choosing l large we get

$$\mathrm{P}\left\{\sup_{x\in(0,T)} \mathrm{D}(x|h_1)^2 > \xi n^{-\frac{4}{9}}\right\} = O\left(n^{-r}\right),\ r > 0,$$

uniformly in $h_1 \in \mathcal{H}_1$. The sum can be bounded as in [16] by using exponential bounds

$$\mathrm{P}\left\{\sup_{x\in[0,T]}\frac{1}{h_2}\sum_{i=1}^{n}\mathrm{I}(|x-X_i|\le h_2)>C_4\right\}\le\mathrm{P}\left\{\sup_{x\in[0,T]}\frac{1}{h_2}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-\frac{|x-X_i|^2}{2h_2^2}}>C_3\right\}=O(n^{-r})$$

for $C_3 \le C_4$ finite constants holds as required and,

$$|b(x|h_1)| = |\mu_2(x|h_1) - \lambda(x)| = \left|\mathbb{E}\hat{\lambda}_2(x|h) - \lambda(x)\right| =$$
$$\left|\lambda(x) + \frac{h_1^2}{2!}\int z^2 K(z)\,dz + \cdots - \lambda(x)\right| \le C_5 h_1^2 \le \left(n^{a-\frac{1}{5}}\right)^2 = C_6 n^{2a-\frac{2}{5}}$$

uniformly in $h_1 \in \mathcal{H}_1$ and $x \in (0,T)$. Combining these last three results we see that the case i=4 is proved.

case i=2: We have,

$$\begin{aligned}\mathbb{E}\left\{\varepsilon_2(x|h_1,h_2)\right\} &= \frac{1}{\hat{h}_2}\mathbb{E}\int\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda(z)^{-\frac{1}{2}}b(z|h_1)}{1-F(z)}\mathrm{L}\left\{\left(\frac{x-z}{\hat{h}_2}\right)\lambda(z)^{\frac{1}{2}}\right\}f(z)\,dz\\ &= \int\lambda(x-h_2z)^{-\frac{1}{2}}b(x-h_2z|h_1)\mathrm{L}\left\{z\lambda(x-h_2z)^{1/2}\right\}\lambda(x-h_2z)\,dz.\end{aligned}$$

Using lemma 3.7.5, we find

$$\mathbb{E}\left\{\varepsilon_2(x|h_1,h_2)\right\} = \int\lambda(x-h_2z)^{\frac{1}{2}}b(x-h_2z|h_1)\mathrm{L}\left\{z\lambda(x-h_2z)^{1/2}\right\}\,dz = O(h_1^2h_2^2)$$

uniformly in $x \in (0,T)$. Thus, working as in the case $i=4$, assuming that l is sufficiently large and using lemma 3.7.6 pp. 66 yields,

$$\begin{aligned}\mathrm{P}\left\{|\varepsilon_2(x)-\mathbb{E}\varepsilon_2(x)|>\xi n^{-4/9}\right\} &\le \left(\xi^{-1}n^{4/9}\right)^l\mathbb{E}\left\{|\varepsilon_2(x)-\mathbb{E}\varepsilon_2(x)|^l\right\}\\ &\le C(l)\xi^{-1}n^{\frac{4l}{9}}\left(\frac{1}{h_2}\right)^l\left\{(h_1^2h_2)^{\frac{l}{2}}+h_1^{2l}h_2^l\right\}\\ &= C(l)\xi^{-1}n^{\frac{4l+1}{9}}\left[n^{\frac{al-23}{90}}+n^{2al-\frac{23l}{45}}\right] = O(n^{-r}).\end{aligned}$$

case i=1: For the last case define

$$m(u,v) = \frac{\lambda(v)^{-\frac{1}{2}}\left\{\frac{K\left(\frac{v-u}{h_1}\right)}{1-F(v)} - h_1\mu_2(v|h_1)\right\}\mathrm{L}\left\{\left(\frac{x-v}{h_2}\right)\lambda(v)^{\frac{1}{2}}\right\}}{1-F(v)},$$
$$m_1(u) = \mathbb{E}\left\{m(u,X_1)\right\} \text{ and } M(u,v) = m(u,v) - m_1(u).$$

Then

$$\sum_{j=1}^{n}m_1(X_j) = \sum_{j=1}^{n}\mathbb{E}\left\{m(X_j,X_1)|X_j\right\} = \sum_{j=1}^{n}t_2(X_j,x|h_1,h_2) = 2nh_1h_2\mathrm{T}_2(x|h_1,h_2).$$

Therefore, (lemma 3.7.7, pp. 67),

$$\varepsilon_1(x|h_1,h_2) = \frac{1}{n^2h_1h_2}\left\{\sum\sum_{i\neq j} M(X_i,X_j) + \sum_{i=1}^{n} M(X_i,X_i)\right\} + 2\mathrm{T}_2(x|h_1,h_2) \tag{3.27}$$

Hence, it remains to show that for $\xi, r > 0$ we have

$$\sup_{x\in[0,T],h_1\in\mathcal{H}_1,h_2\in\mathcal{H}_2} \mathrm{P}\left\{\frac{1}{n^2h_1h_2}\left|\sum_{i=1}^{n} M(X_i,X_i)\right| > \xi n^{-\frac{4}{5}}\right\} = O(n^{-r}) \tag{3.28}$$

$$\sup_{x\in[0,T],h_1\in\mathcal{H}_1,h_2\in\mathcal{H}_2} \mathrm{P}\left\{\frac{1}{n^2h_1h_2}\left|\sum\sum_{i\neq j} M(X_i,X_j)\right| > \xi n^{-\frac{4}{5}}\right\} = O(n^{-r}) \tag{3.29}$$

We treat these two equations separately starting with (3.28).

$$\begin{aligned}\mathbb{E}\sum_{i=1}^{n} M^2(X_i,X_i) \leq \mathbb{E}\sum_{i=1}^{n} m^2(X_i,X_i) &\leq n\int \frac{\lambda(z)^{-1}}{(1-F(z))^4}K^2(0)\mathrm{L}^2\left(\frac{x-z}{h_2}\lambda(z)^{\frac{1}{2}}\right) f(z)\,dz \\ &= \int \frac{nK^2(0)}{(1-F(x))^3}\mathrm{L}^2\left(\frac{x-z}{h_2}\lambda(z)^{\frac{1}{2}}\right) dz \leq C_6 n h_2\end{aligned}$$

with C_6 being a positive generic constant. Also,

$$\sum_{i=1}^{n}\mathbb{E}|M(X_i,X_i)|^l \leq \sum_{i=1}^{n}\mathbb{E}|m(X_i,X_i)|^l = n\int \frac{\lambda(z)^{-\frac{l}{2}}}{(1-F(z))^{2l}}K^l(0)\mathrm{L}^l\left(\frac{x-z}{h_2}\lambda(z)^{\frac{1}{2}}\right) f(z)\,dz$$

and thus,

$$\sum_{i=1}^{n}\mathbb{E}|M(X_i,X_i)|^l \leq nC_7h_2$$

where C_7 is positive generic constant. By (3.26) with $\Phi(x) = x^l$,

$$f^* = \sup_{i=1,\ldots,n}\sum_{i=1}^{n} M(X_i,X_i),\ d_i = M(X_i,X_i) - \mathbb{E}M(X_i,X_i) \text{ and } S(f) = \left(\sum_{i=1}^{n}\mathbb{E}\left\{d_i\right\}^2\right)^{\frac{1}{2}}$$

gives

$$\begin{aligned}\mathbb{E}\left|\sum_{i=1}^{n} M(X_i,X_i)\right|^l &\leq \mathbb{E}\left|\sum_{i=1}^{n}\left\{\sup_{i=1,\ldots,n}\sum_{i=1}^{n} M(X_i,X_i)\right\}\right|^l \\ \leq \left|\sum_{i=1}^{n}\mathbb{E}\left\{M(X_i,X_i)-\mathbb{E}M(X_i,X_i)\right\}^2\right|^{\frac{l}{2}} &+ \sum_{i=1}^{n}\left|\mathbb{E}\left\{M(X_i,X_i)-\mathbb{E}M(X_i,X_i)\right\}\right|^l \\ \leq \left|\sum_{i=1}^{n}\mathbb{E}\left\{M(X_i,X_i)\right\}^2\right|^{\frac{l}{2}} + \sum_{i=1}^{n}\left|\mathbb{E}\left\{M(X_i,X_i)\right\}\right|^l &\leq (n^2h_1h_2)^{-l}((C_6nh_2)^{\frac{l}{2}} + C_7nh_2).\end{aligned}$$

Hence,

$$\mathrm{P}\left\{\frac{1}{n^2h_1h_2}\left|\sum_{i=1}^{n}M(X_i,X_i)\right|>\xi n^{-\frac{4}{5}}\right\}\leq C_7(l)\xi^{-l}\left(\frac{1}{n^2h_1h_2}\right)^{l}$$
$$\times\left\{\mathbb{E}\left|\sum_{i=1}^{n}\left\{M(X_i,X_i)-\mathbb{E}M(X_i,X_i)\right\}^2\right|^{\frac{l}{2}}+\sum_{i=1}^{n}\mathbb{E}\left|M(X_i,X_i)-\mathbb{E}M(X_i,X_i)\right|^{l}\right\}$$
$$\leq C_8(l)\xi^{-l}\left(\frac{1}{n^2h_1h_2}\right)^{l}\left\{(nh_2)^{l/2}+nh_2\right\}.$$

Choosing l sufficiently large completes the proof of (3.28). For the proof of (3.29) first denote with $N(x,y)$ any of the two functions $M(x,y)$, $M(y,x)$. Since the $N(X_i,X_j)$ are identically distributed and since

$$\sum_{i<j}\sum M(X_i,X_j)+\sum_{j<i}\sum M(X_i,X_j)=\sum_{i\neq j}\sum M(X_i,X_j)$$

we can examine only the case where $\sum_{i\neq j}\sum M(X_i,X_j)$ is replaced by $\sum_{i\leq j}\sum N(X_i,X_j)$. Set

$$Z_j=\sum_{1\leq i\leq j-1}N(X_i,X_j)$$

and note that

$$\mathbb{E}\{N(X_i,X_j)|X_i\}=\mathbb{E}\left\{m(X_i,X_j)-m_1(X_i)|X_i\right\}=0$$

for $i\neq j$ so the Z_j's are martingale differences. Also,

$$\sum_{j=1}^{n}\mathbb{E}\{Z_j\}^2=\sum_{j=1}^{n}\mathbb{E}\left\{\sum_{i=1}^{j-1}N(X_i,X_j)\right\}^2=\frac{n(n-1)}{2}\mathbb{E}\left\{N^2(X_1,X_2)\right\}$$

and therefore

$$\mathbb{E}N^2(X_1,X_2)\leq\iint\frac{f(u)}{(1-F(v))^3}K^2\left(\frac{u-v}{h_1}\right)\mathrm{L}^2\left(\frac{x-v}{h_2}\lambda(v)^{\frac{1}{2}}\right)du\,dv\leq C_{10}h_1h_2.$$

Notice that Z_j conditional on X_j is a sum of independent and identically distributed random variables. Thus we can use Rosenthal's inequality to obtain a bound for $\mathbb{E}|Z_j|^l$:

$$\mathbb{E}\left\{|Z_j|^l\right\}\leq C_9\mathbb{E}\left(\left\{(j-1)\mathbb{E}\left\{\left|N^2(X_{j-1},X_j)\right|\Big|X_j\right\}\right\}^{\frac{l}{2}}+(j-1)\mathbb{E}\left\{|N(X_{j-1},X_j)|^l\Big|X_j\right\}\right)$$

and since we have

$$\mathbb{E}\left\{\left|N^2(X_{j-1},X_j)\right|\Big|X_j\right\}\leq\int\frac{1}{(1-F(v))^3}K^2\left(\frac{u-v}{h_1}\right)\mathrm{L}^2\left(\frac{x-v}{h_2}\lambda(v)^{\frac{1}{2}}\right)dv\leq C_{10}h_2$$

and

$$\mathbb{E}\left\{|N(X_{j-1},X_j)|^l\Big|X_j\right\}\leq\int\frac{\lambda^{-\frac{l}{2}}(v)}{(1-F(v))^{2l}}K^l\left(\frac{u-v}{h_1}\right)\mathrm{L}^l\left(\frac{x-v}{h_2}\lambda^{\frac{1}{2}}(v)\right)f(v)\,dv\leq C_{11}h_2$$

we get that

$$\mathbb{E}\left\{|Z_j|^l\right\} \leq C_{12}\left((jh_2)^{\frac{l}{2}} + jh_2\right).$$

Finally from (3.26) with $\Phi(x) = x^l$,

$$f^* = \sup_{i=1,\dots,n} \sum\sum_{1\leq i<j\leq n} N(X_i, X_j),\ d_i = Z_j \ \text{ and } \ S(f) = \left(\sum_{j=1}^{n} \mathbb{E}\left\{Z_j\right\}^2\right)^{\frac{1}{2}}$$

we have that

$$\mathbb{E}\left|\sum\sum_{1\leq i<j\leq n} N(X_i, X_j)\right|^l \leq C_{13}\left((n^2h_1h_2)^{\frac{l}{2}} + n(nh_2)^{\frac{l}{2}}\right)$$

which completes the proof on choosing l large.

3.6.2 Proof of theorem 3.5.2.

Before we give the proof we state and prove the following lemma

Lemma 3.6.1. *The standardized version of S_2, is asymptotically normally distributed with mean zero and variance 1.*

Proof. The proof is based on the Hajek projection method which essentially extents the scope of the central limit theorem to sums that are asymptotically equivalent to sums of independent random variables. Here, we are concerned with the statistic S_2. Applying Hajek's idea we see that the result will follow if we approximate S_2 by its projection, say $\hat{S}$, on the subspace of all such sums of independent terms. Recall the definition of S_2,

$$S_2 = \tilde{\lambda}_{n,2}(x|h) - \mathbb{E}\tilde{\lambda}_{n,2}(x|h).$$

It is immediately seen that it is equivalent to prove that the standardized version of $\tilde{\lambda}_{n,2}$ has asymptotically a standard normal distribution. To prove that we follow the proof of Tanner and Wong, [44]. Write

$$S = \sum_{i=1}^{n} V_i,\quad V_i = \frac{1}{nh}\frac{\lambda(X_i)^{\frac{1}{2}} K\left(\frac{x-X_i}{h}\lambda(X_i)^{\frac{1}{2}}\right)}{1 - F_n(X_i)}$$

and set

$$\hat{S} = \sum_{i=1}^{n} \mathbb{E}(S|X_i) - (n-1)\mathbb{E}S$$

Then we easily see that

$$\mathbb{E}\hat{S} = \mathbb{E}S \ \text{ and } \ \mathbb{E}(S - \hat{S})^2 = \mathbb{V}\mathrm{ar}(S) - \mathbb{V}\mathrm{ar}(\hat{S})$$

Now,

$$\mathbb{E}(V_i|X_i) = \frac{1}{n}Z(X_i) \ \text{ and}$$

$$\mathbb{E}(V_j|X_i) = \frac{1}{n-1}\frac{1}{h}\int \left(1 - F^{n-1}(y)\right) K\left(\frac{x-y}{h}\lambda(y)^{\frac{1}{2}}\right)\lambda(y)^{\frac{3}{2}}\,dy + \frac{1}{n(n-1)}Q(X_i),\quad i \neq j$$

where

$$Z(X_i) = \frac{1-F^n(X_i)}{1-F(X_i)} K\left(\frac{x-X_i}{h}\lambda(X_i)^{\frac{1}{2}}\right)\lambda(X_i)^{\frac{1}{2}}$$
$$Q(X_i) = -\int \frac{1-F^n(y)-nF^{n-1}(y)(1-F(y))}{1-F(y)} K\left(\frac{x-y}{h}\lambda(y)^{\frac{1}{2}}\right)\lambda(y)^{\frac{3}{2}} I_{\{y\le X_i\}}\, dy$$

Now,

$$\begin{aligned}\hat{S}-\mathbb{E}\hat{S} &= \sum_{i=1}^{n}\{\mathbb{E}(V_i|X_i)-(n-1)\mathbb{E}(V_j|X_i)-\mathbb{E}S\}\\ &= \sum_{i=1}^{n}\left\{\frac{1}{n}Z(X_i)+\frac{1}{n}Q(X_i)+R_n\right\}\end{aligned}$$

where

$$R_n = -\int F^{n-1}(y)K\left(\frac{x-y}{h}\lambda(y)^{\frac{1}{2}}\right)\lambda(y)^{\frac{1}{2}} f(y)\, dy$$

It can be easily shown that

$$|Q| = O(\log n), |R_n| = O\left(\frac{1}{n(n+1)}\right) \text{ and } \mathbb{E}|Z(X_i)|^r = \frac{\lambda(x)^{\frac{3}{2}}}{h^{r-1}}\int K^r/(1-F(x))^{r-1}$$

Utilizing these results we can show that $\hat{S}$ and S have the same asymptotic distribution. For that consider

$$\mathbb{V}\mathrm{ar}(\hat{S}) = n\mathbb{V}\mathrm{ar}\left(\frac{1}{n}Z+\frac{1}{n}Q+R_n\right) = \frac{1}{n}\frac{\lambda(x)^{\frac{3}{2}}}{h}\int K^2 + o\left(\frac{1}{nh}\right)$$

therefore, in view of the variance of S_2 we get that $\mathbb{V}\mathrm{ar}(\hat{S})/\mathbb{V}\mathrm{ar}(S_2)\to 1$. Hence,

$$\mathbb{E}\left(\frac{\hat{S}-\mathbb{E}\hat{S}}{\sqrt{\mathbb{V}\mathrm{ar}(\hat{S})}}-\frac{S-\mathbb{E}S}{\sqrt{\mathbb{V}\mathrm{ar}(\hat{S})}}\right)^2 = \frac{\mathbb{E}\left(\hat{S}-S\right)^2}{\mathbb{V}\mathrm{ar}(\hat{S})} = \frac{\mathbb{V}\mathrm{ar}(\hat{S})-\mathbb{V}\mathrm{ar}(S)}{\mathbb{V}\mathrm{ar}(\hat{S})}\to 0.$$

Finally, in order to show that the asymptotic distribution of the standardized statistic $\hat{S}$ is standard normal we use Lyapunov's theorem, [31], pp. 146. According to the theorem and since R_n is negligible, a sufficient condition is that

$$\frac{n\mathbb{E}\left|\frac{1}{n}Z+\frac{1}{n}Q\right|^3}{\sqrt{\mathbb{V}\mathrm{ar}(\hat{S})^3}}\to 0$$

which is already established because using the bounds for Q and Z the above quantity is $O((nh)^{-\frac{1}{2}})$. ∎

Now, we give the proof of theorem 3.5.2

Proof. Substituting $\hat{h}_1, \hat{h}_2$ with non random bandwidth we have that the random variable T_2 is a sum of i.i.d random variables. Thus, asymptotic normality of T_2 may be proved by Linderberg's theorem. Joint asymptotic normality of S_2 and T_2 can then be obtained by using the Cramér-Wold device. Thus we will restrict ourselves to determine asymptotic variances and covariances. From the definition of $m_1(u)$ in page 54,

$$m_1(u)=\mathbb{E}\left\{\frac{\lambda^{-\frac{1}{2}}(X_1)K\left(\frac{u-X_1}{h_1}\right)\mathrm{L}\left(\frac{x-X_1}{h_2}\lambda^{\frac{1}{2}}(X_1)\right)}{(1-F(X_1))^2}\right\}-\mathbb{E}\left\{\frac{\lambda^{-\frac{1}{2}}(X_1)h_1\mu_2(X_1|h_1)\mathrm{L}\left(\frac{x-X_1}{h_2}\lambda^{\frac{1}{2}}(X_1)\right)}{1-F(X_1)}\right\}$$
$$\equiv m_2(u)-m_3.$$

Recall that

$$2nh_1h_2\mathrm{T}_2(x|h_1,h_2)=\sum_{i=1}^{n}m_1(X_i)=\sum_{i=1}^{n}\left\{m_2(X_i)-m_3\right\},$$

where m_3 is a constant and therefore, the asymptotic variance of T_2 is

$$\mathbb{V}\mathrm{ar}\left\{\mathrm{T}_2(x|h_1,h_2)\right\}=\frac{1}{4}\frac{1}{n^2h_1^2h_2^2}\left\{\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}\left\{m_2(X_i)\right\}+2\sum_{i<j}\mathbb{C}\mathrm{ov}\left\{m_2(X_i),m_2(X_j)\right\}\right\}.$$

Note that X_i and X_j are independent and so the covariance is zero. Now,

$$m_2(u)=\int\frac{\lambda^{\frac{1}{2}}(v)}{1-F(v)}K\left(\frac{u-v}{h_1}\right)\mathrm{L}\left(\frac{x-v}{h_2}\lambda^{\frac{1}{2}}(v)\right)\,dv.$$

Let $x-v=h_2z$. Then,

$$m_2(u)=h_2\int\frac{\lambda^{\frac{1}{2}}(x-h_2z)}{1-F(x-h_2z)}K\left(\frac{x-h_2z-u}{h_1}\right)\mathrm{L}\left(z\lambda^{\frac{1}{2}}(x-h_2z)\right)\,dz \tag{3.30}$$

and so

$$\mathbb{E}m_2(X)=h_2\int\left(\int\frac{\lambda^{\frac{1}{2}}(x-h_2z)}{1-F(x-h_2z)}K\left(\frac{x-h_2z-w}{h_1}\right)\mathrm{L}\left(z\lambda^{\frac{1}{2}}(x-h_2z)\right)\,dz\right)f(w)\,dw$$

with the change of variable $x-w=h_1r$ the mean becomes

$$\mathbb{E}m_2(X)=h_2h_1\int\left(\int\frac{\lambda^{\frac{1}{2}}(x-h_2z)}{1-F(x-h_2z)}K\left(r-h_2z\right)\mathrm{L}\left(z\lambda^{\frac{1}{2}}(x-h_2z)\right)\,dz\right)f(x-h_1r)\,dr.$$

Again, using change of variables similar to (3.3)–(3.6) and working as in lemma 3.7.5, the above expression reduces to

$$\mathbb{E}m_2(X)=h_2h_1\lambda(x)\iint K\left(u\right)\mathrm{L}\left(w\right)\,du\,dw=O(h_1h_2),$$

which is negligible and therefore the asymptotic variance for T_2 will follow from

$$\mathbb{V}\mathrm{ar}\left\{\mathrm{T}_2(x|h_1,h_2)\right\}\simeq\frac{1}{4}\frac{1}{n^2h_1^2h_2^2}\sum_{i=1}^{n}\mathbb{E}m_2^2(X_i).$$

Set

$$h = \frac{h_1}{h_2}, \text{ and } v_1 = \frac{z}{h}.$$

Then, $dz = h\,dv_1$ and from (3.30)

$$m_2(u) = h_1 \int \frac{\lambda^{\frac{1}{2}}(x - h_1 v_1)}{1 - F(x - h_1 v_1)} K\left(\frac{x - h_1 v_1 - u}{h_1}\right) \mathrm{L}\left(h v_1 \lambda^{\frac{1}{2}}(x - h_1 v_1)\right) dv_1.$$

From this, we see that

$$\begin{aligned} m_2^2(u) = h_1^2 \iint & \frac{\lambda^{\frac{1}{2}}(x - h_1 v_1)}{1 - F(x - h_1 v_1)} K\left(\frac{x - h_1 v_1 - u}{h_1}\right) \mathrm{L}\left(h v_1 \lambda^{\frac{1}{2}}(x - h_1 v_1)\right) \times \\ & \frac{\lambda^{\frac{1}{2}}(x - h_1 v_2)}{1 - F(x - h_1 v_2)} K\left(\frac{x - h_1 v_2 - u}{h_1}\right) \mathrm{L}\left(h v_2 \lambda^{\frac{1}{2}}(x - h_1 v_2)\right) dv_1\, dv_2. \end{aligned}$$

Therefore

$$\begin{aligned} \mathbb{E}\left\{m_2^2(X)\right\} = h_1^2 \iiint & \frac{\lambda^{\frac{1}{2}}(x - h_1 v_1)}{1 - F(x - h_1 v_1)} K\left(\frac{x - h_1 v_1 - u}{h_1}\right) \mathrm{L}\left(h v_1 \lambda^{\frac{1}{2}}(x - h_1 v_1)\right) \times \\ & \frac{\lambda^{\frac{1}{2}}(x - h_1 v_2)}{1 - F(x - h_1 v_2)} K\left(\frac{x - h_1 v_2 - u}{h_1}\right) \mathrm{L}\left(h v_2 \lambda^{\frac{1}{2}}(x - h_1 v_2)\right) f(u)\, dv_1\, dv_2\, du. \end{aligned}$$

Set

$$w = \frac{x - h_1 v_1 - u}{h_1} \Leftrightarrow \frac{x - u}{h_1} = w + v_1,$$

and note that

$$\frac{x - h_1 v_2 - u}{h_1} = \frac{x - u}{h_1} - v_2 = w + v_1 - v_2.$$

Then,

$$\begin{aligned} \mathbb{E}\left\{m_2^2(X)\right\} = h_1^3 \iiint & \frac{\lambda^{\frac{1}{2}}(x - h_1 v_1)}{1 - F(x - h_1 v_1)} K(w) \mathrm{L}\left(h v_1 \lambda^{\frac{1}{2}}(x - h_1 v_1)\right) \frac{\lambda^{\frac{1}{2}}(x - h_1 v_2)}{1 - F(x - h_1 v_2)} \\ & \times K(w + v_1 - v_2) \mathrm{L}\left(h v_2 \lambda^{\frac{1}{2}}(x - h_1 v_2)\right) f(x - h_1 v_1 - h_1 w)\, dv_1\, dv_2\, dw. \end{aligned}$$

Put

$$Q(x - h_1 s) = \frac{\lambda^{\frac{1}{2}}(x - h_1 s)}{1 - F(x - h_1 s)}.$$

Expanding Q and f in Taylor series around x and setting $a = \lambda^{1/2}(\cdot)$ gives

$$\begin{aligned} \mathbb{E} m_2^2(X) &\simeq h_1^3 Q^2(x) f(x) \iiint K(v) \mathrm{L}(a h v_1) K(v + v_1 - v_2) \mathrm{L}(a h v_2)\, dv\, dv_1\, dv_2 \\ &= h_1^3 Q^2(x) f(x) \iiint K(v) L(a h (v_2 + z)) K(v + z) L(a h v_2)\, dv\, dv_2\, dz \end{aligned}$$

Replacing $\int K(v)K(v+z)\,dv$ with $M(z)$ and $hv_2 = u$,

$$\iiint K(v)\mathrm{L}\left\{ah(v_2+z)\right\} K(v+z)\mathrm{L}(ahv_2)\,dv\,dv_2\,dz =$$

$$\frac{1}{h}\iint \mathrm{L}(au+ahz)M(z)L(au)\,du\,dz \simeq \frac{1}{h}\iint \mathrm{L}^2(au)M(z)\,du\,dz = \frac{a^{-1}}{h}\int \mathrm{L}^2(u)\,du.$$

Hence,

$$\mathbb{E}m_2^2(X) \simeq \frac{h_1^3}{h}\frac{\lambda(x)}{(1-F(x))^2}f(x)\lambda^{-\frac{1}{2}}(x)\int \mathrm{L}^2(u)\,du = h_1^2h_2\frac{\lambda^{\frac{3}{2}}(x)}{1-F(x)}\int \mathrm{L}^2(u)\,du.$$

The covariance will be calculated from

$$\mathbb{C}\mathrm{ov}\left(S_2(x|h_2), \mathrm{T}_2(x|h_1,h_2)\right) = \mathbb{E}\left\{S_2(x|h_2)\mathrm{T}_2(x|h_1,h_2)\right\} - \mathbb{E}\left\{S_2(x|h_2)\right\}\mathbb{E}\left\{\mathrm{T}_2(x|h_1,h_2)\right\}.$$

Since $\mathbb{E}\left\{\mathrm{T}_2(x|h_1,h_2)\right\} = 0$ and $\mathbb{E}\left\{S_2(x|h_2)\right\} = 0$ the above formula extends to

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left(S_2(x|h_2), \mathrm{T}_2(x|h_1,h_2)\right) &= \mathbb{E}\left\{S_2(x|h_2)\mathrm{T}_2(x|h_1,h_2)\right\} \\ &= \mathbb{E}\left\{\left(\tilde{\lambda}_{n,2}(x|h_2) - \mathbb{E}\tilde{\lambda}_{n,2}(x|h_2)\right)\mathrm{T}_2(x|h_1,h_2)\right\}. \qquad (3.31)\end{aligned}$$

Now, by considering only the leading term of the product of the means of $\tilde{\lambda}_{n,2}$ and m_2 we have,

$$\left\{\mathbb{E}\tilde{\lambda}_{n,2}(x|h_2)\right\}\left\{\mathbb{E}m_2(X_i)\right\} \simeq \lambda^2(x)h_2h_1\iint K(u)\,\mathrm{L}(w)\;du\,dw,$$

and thus

$$\mathbb{E}\tilde{\lambda}_{n,2}(x|h_2)\mathbb{E}\mathrm{T}_2(x|h_1,h_2) \simeq \frac{\lambda^2(x)}{2}\iint K(u)\,\mathrm{L}(w)\;du\,dw. \qquad (3.32)$$

To compute $\mathbb{E}\left\{\tilde{\lambda}_n(x|h_2)T(x|h_1,h_2)\right\}$, separate the diagonal and non diagonal terms. For the diagonal term note that

$$\begin{aligned}&\frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^n m_2(X_i)\lambda(X_i)^{1/2}\frac{K\left(\frac{x-X_i}{h_2}\lambda^{1/2}(X_i)\right)}{1-F_n(X_i)}\right\} \\ &\qquad = \mathbb{E}\left\{\sum_{i=1}^n m_2(X_{(i)})\frac{\lambda^{1/2}(X_{(i)})K\left(\frac{x-X_{(i)}}{h_2}\lambda^{1/2}(X_{(i)})\right)}{n-i+1}\right\}. \qquad (3.33)\end{aligned}$$

Let $f_{X_{(i)}}(u)$ denote the density of the i-th order statistic, then

$$
\begin{aligned}
&\mathbb{E}\Big\{\sum_{i=1}^{n} m_2(X_{(i)})\frac{\lambda^{1/2}(X_{(i)})K\big(\frac{x-X_{(i)}}{h_2}\lambda^{1/2}(X_{(i)})\big)}{n-i+1}\Big\}\\
&=\sum_{i=1}^{n}\mathbb{E}\Big\{\Big[\int\frac{\lambda^{1/2}(v)}{1-F(v)}K(\frac{X_{(i)}-v}{h_1})L\left(\frac{x-v}{h_2}\lambda^{1/2}(v)\right)dv\Big]\times\\
&\qquad\qquad\frac{\lambda^{1/2}(X_{(i)})K\big(\frac{x-X_{(i)}}{h_2}\lambda^{1/2}(X_{(i)})\big)}{n-i+1}\Big\}\\
&=\sum_{i=1}^{n}\iint Q(v)K\left(\frac{u-v}{h_1}\right)L\left(\frac{x-v}{h_2}\lambda^{1/2}(v)\right)\times\\
&\qquad\qquad\frac{\lambda^{1/2}(u)K\left(\frac{x-u}{h_2}\lambda^{1/2}(u)\right)}{n-i+1}f_{X_{(i)}}(u)dudv\\
&=\iint Q(v)K\left(\frac{u-v}{h_1}\right)L\left(\frac{x-v}{h_2}\lambda^{1/2}(v)\right)\lambda^{1/2}(u)K\left(\frac{x-u}{h_2}\lambda^{1/2}(u)\right)\times\\
&\qquad\qquad\lambda(u)(1-F^n(u))dudv\\
&\simeq\iint Q(v)K\left(\frac{u-v}{h_1}\right)L\left(\frac{x-v}{h_2}\lambda^{1/2}(v)\right)\lambda^{3/2}(u)K\left(\frac{x-u}{h_2}\lambda^{1/2}(u)\right)dudv\\
&=I\,.
\end{aligned}
$$

Now set $v=x-h_1y$, $u=x-h_2w$ and $hh_2=h_1$. Then,

$$
\begin{aligned}
I=h_1h_2\iint &Q(x-h_1y)K\left(\frac{hy-w}{h}\right)L\big(hy\lambda^{1/2}(x-h_1y)\big)\times\\
&\lambda^{3/2}(x-h_2w)K\big(w\lambda^{1/2}(x-h_2w)\big)dydw\,.
\end{aligned}
$$

Next use substitution $hy-w=hu$ and then, expanding Q and λ in Taylor series around x, note that

$$I\simeq h_1h_2Q(x)\lambda^{3/2}(x)\iint K(u)L(aw)K(aw)dudw,$$

where a is as defined before. Finally set $aw=v$ to conclude that

$$I\simeq h_1h_2\frac{\lambda^{3/2}(x)}{1-F(x)}\int K(v)L(v)dv. \tag{3.34}$$

For the non diagonal term, with similar arguments and algebra, but using the joint density of $(X_{(i)},X_{(j)})$, one can derive that

$$
\begin{aligned}
&\frac{1}{2n^2h_1h_2^2}\mathbb{E}\left\{\sum\sum_{i\neq j}m_2(X_j)\lambda^{1/2}(X_i)\frac{K\left(\frac{x-X_i}{h_2}\lambda^{1/2}(X_i)\right)}{1-F_n(X_i)}\right\}\simeq\\
&\qquad\qquad\frac{\lambda^2(x)}{2}\iint K(u)L(w)\,du\,dw.
\end{aligned}\tag{3.35}
$$

By combining (3.32) – (3.35), and substituting back to (3.31) we find

$$\mathbb{C}\mathrm{ov}\left(S_2(x|h_2), \mathrm{T}_2(x|h_1, h_2)\right) \simeq \frac{1}{2}\frac{\lambda^{\frac{3}{2}}(x)}{1 - F(x)} \int K(u)\mathrm{L}(u)\, du.$$

3.7 Lemmas.

Lemma 3.7.1 (Equation (3.19)).

$$\lambda(x)^{\frac{1}{2}}\left[1 + \frac{\mathrm{D}(x|\hat{h}_1) + b(x|\hat{h}_1)}{\lambda(x)}\right]^{\frac{1}{2}} = \lambda(x)^{\frac{1}{2}}\left[1 + \frac{\hat{\lambda}_2(x|\hat{h}_1) - \mu_2(x|\hat{h}_1) + \mu_2(x|\hat{h}_1) - \lambda(x)}{\lambda(x)}\right]^{\frac{1}{2}}$$
$$= \lambda(x)^{\frac{1}{2}}\left[1 + \frac{\hat{\lambda}_2(x|\hat{h}_1)}{\lambda(x)} - 1\right]^{\frac{1}{2}} = \lambda(x)^{\frac{1}{2}}\frac{\hat{\lambda}_2(x|\hat{h}_1)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}} = \hat{\lambda}_2(x|\hat{h}_1)^{\frac{1}{2}}.$$

Lemma 3.7.2 (Equation (3.22)).

Proof. In order to show that (3.22) is true we only need to show that

$$\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}\delta(X_i)\mathrm{L}\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1 - F(X_i)} + \delta_3(x) = \frac{1}{2}\varepsilon_1(x|\hat{h}_1, \hat{h}_2) + \frac{1}{2}\varepsilon_2(x|\hat{h}_1, \hat{h}_2)$$
$$+ \varepsilon_3(x|\hat{h}_1, \hat{h}_2).$$

Multiply and divide the numerator and the denominator of both ε_1 and ε_2 by $\lambda(X_i)$. Then,

$$\frac{\varepsilon_1(x|\hat{h}_1, \hat{h}_2) + \varepsilon_2(x|\hat{h}_1, \hat{h}_2)}{2} = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda^{\frac{1}{2}}(X_i)\delta_4(X_i)\mathrm{L}\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1 - F(X_i)}.$$

Since

$$\sup_{x\in[0,T]}\{\mathrm{D}(x|\hat{h}_1)^2 + b(x|\hat{h}_1)^2\} \leq \sup_{x\in[0,T]}\left\{\mathrm{D}(x|\hat{h}_1) + b(x|\hat{h}_1)\right\}^2,$$

taking

$$\varepsilon_4(x|\hat{h}_1, \hat{h}_2) = \frac{1}{n\hat{h}_2}\left\{\sup_{x\in[0,T]}\{\mathrm{D}(x|\hat{h}_1)^2 + b(x|\hat{h}_1)^2\}\sum_{i=1}^{n}\mathrm{I}\left(|x - X_i| \leq C_2\hat{h}_2\right)\right\}$$

(3.22) follows immediately.

Lemma 3.7.3 (Equivalence of (3.23) with (3.24) and (3.25)).

Proof. Applying the definition of convergence in probability to (3.23) we see that we need to show that

$$\mathrm{P}\left\{\sup_{x\in[0,T]}\left|\frac{1}{2}\varepsilon_1(x|h_1,h_2)+\frac{1}{2}\varepsilon_2(x|h_1,h_2)+\varepsilon_3(x|h_1,h_2)-\mathrm{T}_2(x|h_1,h_2)\right|>\xi n^{-\frac{4}{9}}\right\}\xrightarrow{p}0.$$

Now,

$$\mathrm{P}\left\{\sup_{x\in[0,T]}\left|\frac{\varepsilon_1(x|h_1,h_2)+\varepsilon_2(x|h_1,h_2)}{2}+\varepsilon_3(x|h_1,h_2)-\mathrm{T}_2(x|h_1,h_2)\right|>\xi n^{-\frac{4}{9}}\right\}$$
$$\leq \mathrm{P}\left\{\sup_{x\in[0,T]}\left|\frac{1}{2}\varepsilon_1(x|h_1,h_2)-\mathrm{T}_2(x|h_1,h_2)\right|>\frac{\xi n^{-\frac{4}{9}}}{3}\right\}+$$

$$+\mathrm{P}\left\{\sup_{x\in[0,T]}\left|\frac{1}{2}\varepsilon_2(x|h_1,h_2)\right|>\frac{\xi n^{-\frac{4}{9}}}{3}\right\}+\mathrm{P}\left\{\sup_{x\in[0,T]}|\varepsilon_3(x|h_1,h_2)|>\frac{\xi n^{-\frac{4}{9}}}{3}\right\}$$

In order to show that every term in the above sum is $O\left(n^{-r}\right)$, for every n we consider a finite subset $\mathcal{X}$ of $(0,T)$. Then for all $x\in\mathcal{X}$ there exists $y\in[0,T]$ such that $|x-y|<n^{-s}$, for an arbitrary positive s. Since the functions ε_i, $i=1,2,4$ are Hölder continuous, as sums of Hölder continuous functions, the desired will be true if we show that for any $\xi,r>0$

$$\sup_{x\in[0,T]}\mathrm{P}\left\{\left|\frac{1}{2}\varepsilon_1(x|h_1,h_2)-\mathrm{T}_2(x|h_1,h_2)\right|>\xi n^{-\frac{4}{9}}\right\}=O\left(n^{-r}\right)$$
$$\sup_{x\in[0,T]}\mathrm{P}\left\{\left|\frac{1}{2}\varepsilon_i(x|h_1,h_2)\right|>\xi n^{-\frac{4}{9}}\right\}=O\left(n^{-r}\right),\quad i=2,4. \qquad \blacksquare$$

Lemma 3.7.4 (Bound for $|\mathrm{D}(x|h_1)|$).

Proof. Define the sequence of σ-fields $\mathcal{F}_n=\sigma(X_1,X_2,\dots,X_n)$. Set

$$\lambda_i(x)=\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}-\mathbb{E}\left\{\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}\right\}.$$

For every $i=1,2,\dots,n$

$$\mathbb{E}\lambda_i(x)=\mathbb{E}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}-\mathbb{E}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}=0$$

and

$$\mathbb{E}\left(\sum_{i=1}^{n}\lambda_i(x)|\mathcal{F}_{n-1}\right)=\mathbb{E}\left(\sum_{i=1}^{n-1}\lambda_i(x)|\mathcal{F}_{n-1}\right)+\mathbb{E}\left(\lambda_n(x)|\mathcal{F}_{n-1}\right)=\sum_{i=1}^{n-1}\lambda_i(x),$$

i.e. $\sum_{i=1}^{n} \lambda_i(x)$ is a martingale with respect to the sequence of σ-fields generated by $\{\mathcal{F}_n, n \geq 1\}$. Since $\mathrm{D}(x|h_1) - \bar{\mathrm{D}}(x|h_1) \to 0$ a.s. as $n \to +\infty$, where

$$\bar{\mathrm{D}}(x|h_1) = \bar{\lambda}_2(x|h_1) - \mathbb{E}\left\{\bar{\lambda}_2(x|h_1)\right\}$$

we replace D by $\bar{\mathrm{D}}$.

$$\bar{\mathrm{D}}(x|h_1) = \bar{\lambda}_2(x|h_1) - \mathbb{E}\bar{\lambda}_2(x|h_1) = \frac{1}{nh_1}\sum_{i=1}^{n}\left\{\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)} - \mathbb{E}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}\right\} = \frac{1}{nh_1}\sum_{i=1}^{n}\lambda_i(x).$$

Now,

$$\mathbb{E}\sum_{i=1}^{n}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2 \leq \mathbb{E}\sum_{i=1}^{n}\left\{\frac{1}{nh_1}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}\right\}^2 = \frac{1}{nh_1^2}\int K^2\left(\frac{x-y}{h_1}\right)\frac{\lambda(y)}{1-F(y)}\,dy.$$

Set

$$\frac{x-y}{h_1} = z \Rightarrow \; dy = -h_1\,dz.$$

Then,

$$\mathbb{E}\sum_{i=1}^{n}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2 \leq \frac{1}{nh_1}\int K^2(z)\frac{\lambda(x-h_1z)}{1-F(x-h_1z)}\,dz.$$

Approximating $\lambda(x-h_1z)/(1-F(x-h_1z))$ with its Taylor expansion around x, and since by conditions A1 and A3 $R(K)$ is finite, there is a constant $C_1 > 0$ such that

$$\mathbb{E}\sum_{i=1}^{n}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2 \leq \frac{C_1}{nh_1}.$$

Also,

$$\sum_{i=1}^{n}\mathbb{E}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^l \leq \mathbb{E}\left\{\frac{1}{nh_1}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}\right\}^l = \frac{1}{n^{l-1}h_1^l}\int K^l\left(\frac{x-y}{h_1}\right)\frac{\lambda(y)}{(1-F(y))^l}\,dy.$$

Using the same change of variable and Taylor expansion as above, there is a constant $C_2 > 0$ such that

$$\mathbb{E}\sum_{i=1}^{n}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^l \leq \frac{C_2}{(nh_1)^{l-1}}.$$

Then, a bound for $\mathbb{E}|\mathrm{D}(x|h_1)|^l$ can be found by (3.26) with

$$\Phi(x) = x^l,\; f^* = \sup_{i=1,\dots,n}\sum_{i=1}^{n}\frac{\lambda_i(x)}{nh_1},\; d_i = \frac{\lambda_i(x)}{nh_1},\; \text{and } S(f) = \left(\sum_{i=1}^{n}\mathbb{E}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2\right)^{\frac{1}{2}}.$$

Applying (3.26) yields,

$$\mathbb{E}\left|\sup_{i=1,\dots,n}\sum_{i=1}^{n}\frac{\lambda_i(x)}{nh_1}\right|^l \leq C\left[\sum_{i=1}^{n}\mathbb{E}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2\right]^{\frac{l}{2}} + c\sum_{i=1}^{n}\mathbb{E}\left|\frac{\lambda_i(x)}{nh_1}\right|^l.$$

Since,

$$\mathbb{E}|\bar{\mathrm{D}}(x|h_1)|^l \leq \mathbb{E}\left|\sup_{i=1,\ldots,n}\sum_{i=1}^{n}\frac{\lambda_i(x)}{nh_1}\right|^l$$

we finally get that

$$\mathbb{E}|\bar{\mathrm{D}}(x|h_1)|^l \leq C_3\left\{\left(\frac{1}{nh_1}\right)^{\frac{l}{2}} + \frac{1}{(nh_1)^{l-1}}\right\}. \qquad \blacksquare$$

Lemma 3.7.5 (bound for $\varepsilon_2(x|h_1,h_2)$).

Proof.

$$\mathbb{E}\left\{\varepsilon_2(x|h_1,h_2)\right\} = \int \lambda^{\frac{1}{2}}(x-h_2z)b(x-h_2z|h_1)\mathrm{L}\left\{z\lambda^{\frac{1}{2}}(x-h_2z)\right\}\,dz. \tag{3.36}$$

Setting as before

$$\begin{aligned}
&u(z) = \frac{\lambda^{\frac{1}{2}}(x-z)}{\lambda^{\frac{1}{2}}(x)} \Rightarrow u(h_2z) = \frac{\lambda^{\frac{1}{2}}(x-h_2z)}{\lambda^{\frac{1}{2}}(x)},\\
&z\lambda^{\frac{1}{2}}(x) = y \Rightarrow dy = \lambda^{\frac{1}{2}}(x)dz,\\
&\text{and } \ \eta = \frac{h_2}{\lambda^{\frac{1}{2}}(x)}
\end{aligned}$$

yields

$$u(h_2z) = u\left(\frac{h_2y}{\lambda^{\frac{1}{2}}(x)}\right) = u(\eta y).$$

Then,

$$\begin{aligned}
\mathbb{E}\left\{\varepsilon_2(x|h_1,h_2)\right\} &= \int \frac{\lambda^{\frac{1}{2}}(x)}{\lambda^{\frac{1}{2}}(x)}\lambda^{\frac{1}{2}}(x-h_2z)b(x-h_2z|h_1)\mathrm{L}\left\{z\frac{\lambda^{\frac{1}{2}}(x)}{\lambda^{\frac{1}{2}}(x)}\lambda^{\frac{1}{2}}(x-h_2z)\right\}\,dz\\
&= \int u(\eta y)b(\eta y|h_1)\mathrm{L}\left\{yu(\eta y)\right\}\,dy. \qquad (3.37)
\end{aligned}$$

Expanding b in Taylor series around 0 and using (3.7) in L and expanding in Taylor series, i.e.

$$b(\eta y) = b(0) + b'(0)\eta y + \frac{(\eta y)^2}{2!}b''(0) + \ldots$$

and

$$\begin{aligned}
\mathrm{L}\left(yu(\eta y)\right) &= \mathrm{L}\left\{y\left(u(0) + u'(0)\eta y + \frac{(\eta y)^2}{2!}u''(0) + \ldots\right)\right\}\\
&= \mathrm{L}\left(yu(0) + u'(0)\eta y^2 + \frac{\eta^2y^3}{2!}u''(0) + \ldots\right).
\end{aligned}$$

Then,

$$\mathrm{L}\left(yu(\eta y)\right) = \mathrm{L}(y) + \left(\eta y^2 u'(0) + \frac{\eta^2 y^3}{2!}u'(0) + \dots\right)\mathrm{L}'(y)$$
$$+ \left(\eta y^2 u'(0) + \frac{\eta^2 y^3}{2!}u'(0) + \dots\right)\mathrm{L}''(y) + \dots$$

and substituting to (3.37) gives

$$\mathbb{E}\left\{\varepsilon_2(x|h_1, h_2)\right\} = \int \left\{\mathrm{L}(y) + \eta\left(\mathrm{L}(y)u'(0)y + yb'(0)\mathrm{L}(y) + y^2u'(0)\mathrm{L}'(y)\right) + \dots\right\} dy.$$

Note that

$$\int \mathrm{L}(y) = \int \left\{K(y) + yK'(y)\right\} dy = \int K(y)\,dy + yK(y) - \int K(y)\,dy = 0 \quad (3.38)$$

and $$\int y\mathrm{L}(y) = \int yK(y)\,dy + \int y^2K'(y)\,dy = 0. \quad (3.39)$$

Rearranging, and since $b(x|h_1) \le C_1h_1^2$ and $\lambda(x - h_2z)/\lambda(x) \le C_2h_2^2$ we get finally that

$$\mathbb{E}\left\{\varepsilon_2(x|h_1, h_2)\right\} \le Mh_1^2h_2^2$$

where M is a positive generic constant. ∎

Lemma 3.7.6 (Bound for $\mathbb{E}|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)|$).

Proof. Set

$$r_i(x) = \frac{1}{nh_2}\frac{\lambda^{-\frac{1}{2}}(X_i)b(X_i|h_1)\mathrm{L}\left\{\left(\frac{x-X_i}{h_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1 - F(X_i)}.$$

Then

$$\mathbb{E}|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)| = \mathbb{E}\left|\sum_{i=1}^n r_i(x) - \mathbb{E}\sum_{i=1}^n r_i(x)\right| = \mathbb{E}\left|\sum_{i=1}^n\left\{r_i(x) - \mathbb{E}r_i(x)\right\}\right|.$$

Now, let

$$Z_i = r_i(x) - \mathbb{E}r_i(x), \quad i = 1, \dots, n.$$

For every $i = 1, 2, \dots, n$

$$\mathbb{E}Z_i = \mathbb{E}r_i(x) - \mathbb{E}r_i(x) = 0.$$

Also,

$$\mathbb{E}\left(\sum_{i=1}^n Z_i|\mathcal{F}_{n-1}\right) = \mathbb{E}\left(\sum_{i=1}^{n-1} Z_i|\mathcal{F}_{n-1}\right) + \mathbb{E}\left(Z_n|\mathcal{F}_{n-1}\right) = \sum_{i=1}^{n-1} Z_i,$$

i.e. $\sum_{i=1}^n Z_i$ is a martingale with respect to the sequence of σ-fields generated by $\{\mathcal{F}_n, n \ge 1\}$. Now,

$$\sum_{i=1}^n \mathbb{E}(Z_i)^2 \le \sum_{i=1}^n \mathbb{E}(r_i(x))^2 = \int \sum_{i=1}^n \frac{1}{n^2h_2^2}\frac{\lambda(u)b^2(u|h_1)}{(1 - F(u))^2}\mathrm{L}^2\left(\frac{x-u}{h_2}\lambda^{\frac{1}{2}}(u)\right) f(u)\,du$$
$$= \int \frac{1}{nh_2^2}\frac{\lambda^2(u)}{1 - F(u)}b^2(u|h_1)\mathrm{L}^2\left(\frac{x-u}{h_2}\lambda^{\frac{1}{2}}(u)\right) du.$$

Since $|b(x|h_1)| \leq C_6 h_1^2$, we have that $b(x|h_1)^2 \leq C_6^2 h_1^4$. Also, from conditions A1, A3, $R(K + zK')$ is finite. Thus, using changes of variables similar to those used in the proof of theorem 3.3.2 for the variance,

$$\sum_{i=1}^{n} \mathbb{E}(Z_i)^2 \leq C_7 \frac{h_1^4}{nh_2}.$$

Also,

$$\sum_{i=1}^{n} \mathbb{E}\{Z_i\}^l \leq \sum_{i=1}^{n} \int \frac{1}{n^l h_2^l} \frac{\lambda^{\frac{l}{2}}(u) b^l(u|h_1)}{(1-F(u))^l} \mathrm{L}^l \left(\frac{x-u}{h_2} \lambda^{\frac{1}{2}}(u) \right) f(u)\, du \leq \frac{C_7 h_1^{2l}}{(nh_2)^{l-1}}.$$

Using (3.26) with

$$\Phi(x) = x^l,\ f^* = \sup_{i=1,\dots,n} \sum_{i=1}^{n} Z_i,\ d_i = Z_i, \text{ and } S(f) = \left(\sum_{i=1}^{n} \mathbb{E}\left\{Z_i\right\}^2 \right)^{\frac{1}{2}}$$

gives

$$\mathbb{E}\left| \sup_{i=1,\dots,n} \sum_{i=1}^{n} Z_i \right|^l \leq C \left[\sum_{i=1}^{n} \mathbb{E}\left\{Z_i\right\}^2 \right]^{\frac{l}{2}}$$

and therefore

$$\mathbb{E}|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)| \leq C_8 \left\{ \frac{h_1^{2l}}{(nh_2)^{\frac{l}{2}}} + \frac{h_1^{2l}}{(nh_2)^{l-1}} \right\} = C_8 \frac{h_1^{2l}(nh_2)^{l-1} + h_1^{2l}(nh_2)^{\frac{l}{2}}}{(nh_2)^{\frac{3l}{2}-1}}.$$

Recall that $h_2 \geq \eta n^{-1/9}$.

$$\frac{1}{h_2} \leq \frac{1}{\eta n^{-\frac{l}{9}}} \Rightarrow \frac{1}{nh_2} \leq \frac{1}{\eta n^{\frac{8}{9}}} \Rightarrow \left(\frac{1}{nh_2} \right)^{\frac{3l}{2}-1} \leq \frac{1}{\eta^{\frac{3l}{2}-1} n^{\frac{4l}{3}-\frac{8}{9}}}$$

Also

$$h_1^{2l} \leq \rho n^{2al - \frac{2l}{5}},\ \ (nh_2)^{l-1} \leq n^{-\frac{l}{9}}$$

and therefore

$$h_1^{2l}(nh_2)^{l-1} \leq C n^{-\frac{l}{9}} n^{2al-\frac{2l}{5}} = C n^{2al - \frac{19l}{45}}$$

Similarly

$$h_1^{2l}(nh_2)^{\frac{l}{2}} \leq n^{-\frac{l}{9}} \left(n\rho n^{a-\frac{2}{5}} \right)^{\frac{l}{2}}.$$

Hence,

$$\mathbb{E}|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)| \leq C \frac{n^{2al-\frac{19l}{45}} + n^{al+\frac{3l}{10}-\frac{l}{9}}}{\eta^{\frac{3l}{2}-1} n^{\frac{4l}{3}}}. \quad \blacksquare$$

Lemma 3.7.7 (Equation (3.27)).

Proof. From the definition of the function $M(x,y)$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} M(X_{(i)},X_{(j)}) = \sum\sum_{i\neq j} M(X_i,X_j) + \sum_{i=1}^{n} M(X_i,X_i).$$

Thus,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} M(X_{(i)},X_{(j)}) = \sum\sum_{i\neq j}\left\{m\left(X_i,X_j\right) - m_1\left(X_i\right)\right\} + \sum_{i=1}^{n}\left\{m\left(X_i,X_i\right) - m_1\left(X_i\right)\right\}$$

$$= \sum\sum_{i\neq j} m\left(X_i,X_j\right) + \sum_{i=1}^{n} m\left(X_i,X_i\right) - \left((n-1)\sum_{i=1}^{n} m_1\left(X_i\right) + \sum_{i=1}^{n} m_1\left(X_i\right)\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} m\left(X_i,X_j\right) - n\sum_{i=1}^{n} m_1\left(X_i\right).$$

Therefore,

$$\frac{1}{n^2h_1h_2}\left\{\sum\sum_{i\neq j} M(X_i,X_j) + \sum_{i=1}^{n} M(X_i,X_i)\right\} + 2\mathrm{T}_2(x|h_1,h_2) =$$

$$\frac{1}{n^2h_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} m\left(X_i,X_j\right) - \frac{1}{nh_1h_2}\sum_{i=1}^{n} m_1\left(X_i\right) + \frac{1}{nh_1h_2}\sum_{i=1}^{n} m_1\left(X_i\right)$$

$$= \frac{1}{n^2h_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} m\left(X_i,X_j\right).$$

Now,

$$\frac{1}{n^2h_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} m(X_i,X_j) =$$

$$\frac{1}{n^2h_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\lambda(X_j)^{\frac{1}{2}}h_1\left\{\frac{K_{h_1}(X_j-X_i)}{1-F(X_j)} - \mu(X_j|h_1)\right\}\mathrm{L}\left\{\frac{x-X_j}{h_2}\lambda(X_j)^{\frac{1}{2}}\right\}}{1-F(X_j)} =$$

$$\frac{1}{nh_2}\sum_{i=1}^{n}\frac{\lambda^{-\frac{1}{2}}(X_i)\mathrm{D}(X_i|h_1)\mathrm{L}\left\{\left(\frac{x-X_i}{h_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)} = \varepsilon_1(x|h_1,h_2). \quad \blacksquare$$

Chapter 4

TRANSFORMATIONS IN FAILURE RATE ESTIMATION

4.1 Introduction.

The variable bandwidth estimators studied in the previous chapter reduce the bias compared to usual kernel estimators defined in the introduction and at the same time avoid the undesirable consequences generated by using estimators based on higher order kernels. In this chapter we present another method, based on transformations of random variables which is in the same spirit with the variable bandwidth approach as far as bias reduction is concerned. That is, the resulting estimators perform visibly better than the usual second order kernel estimates for finite samples and at the same time avert the adverse side effects of higher order kernel-based estimates.

The basic idea here is that there will not be any bias (except for short intervals on either ends) if we were to use our nonparametric estimator to estimate a constant hazard rate function. Making use of this fact we motivate our transformation-based estimator in the next section. We restrict our analytical study to outline the error properties of the proposed estimator which is presented in section 4.3. In section 4.4 we discuss implementation of variable bandwidth adaptive estimators of the last chapter and transformation-based estimators in the present chapter.

4.2 Motivation.

Let $X_1, \ldots X_n$ be independent observations from some density $f_X(x)$ having cdf $F_X(x)$. λ_X denotes the true hazard rate and Λ_X the integrated hazard rate, i.e.

$$\begin{aligned}
\lambda_X(x) &= \frac{f_X(x)}{1 - F_X(x)}, \\
\Lambda_X(x) &= \int_{-\infty}^{x} \lambda_X(u)\, du. \qquad (4.1)
\end{aligned}$$

Recall from the first chapter the definition of estimator $\hat{\lambda}_2(x|h)$, which throughout this chapter we denote as $\hat{\lambda}_{T,1}$,

$$\hat{\lambda}_2(x|h) \equiv \hat{\lambda}_{T,1}(x|h) = \sum_{i=1}^{n} \frac{K_h\left(x - X_{(i)}\right)}{n - i + 1}.$$

Provided that a symmetric kernel K is used and that λ is sufficiently smooth, the asymptotic bias of this estimator has a formal expansion of the form

$$\mathbb{E}\hat{\lambda}_{T,1}(x|h) - \lambda(x) = \sum_{i=1}^{+\infty} \frac{h^{2i}}{(2i)!}\lambda^{(2i)}(x) \int u^{2i} K(u)\, du + o(n^{-1}).$$

A higher order kernel-based hazard rate estimate, utilizes the fact that the first $m(> 2)$ moments of the kernel, for some even integer m, are all zero. Then, as it can be seen from the above formula, the bias of the estimator can be reduced to any desired order.

The transformations method is based on the observation that when an estimator such as $\hat{\lambda}_{T,1}$ is used, bias is completely eliminated when $\lambda^{(2i)}(x) = 0$ for every i. In particular, this is true if the $X_i's$ are exponentially distributed as in this case, the hazard rate function is constant.

The method takes advantage of this property by transforming the original sample $X_1, \dots, X_n$ to the exponential sample $Y_i = g(X_i)$, $i = 1, \dots, n$, where g is a smooth monotonic function. Then, by estimating the hazard rate function of the transformed data and transforming back the estimate, gives the estimate of the hazard rate function of interest. The first of the following two theorems provides the function g which will transform the original sample to a sample from exponential distribution. The second theorem, stated without proof, provides the formula for the hazard rate of a transformed random variable in terms of the hazard rate of the original random variable and the transform-function.

Theorem 4.2.1. *Let Y be a continuous positive random variable with absolutely continuous distribution function F. Let $\Lambda_{\mathrm{X}}(X)$ be as in* (4.1)*. Then the random variable $Y = \Lambda_{\mathrm{X}}(X)$ has standard exponential distribution.*

Proof. This is a consequence of the fact that the r.v. $Z = F(X)$ has $U[0,1]$ distribution. To see that observe that for $z < 0$, $F(z) = 0$ and for $z \geq 0$ write

$$\mathrm{P}\{Z \leq z\} = \mathrm{P}\{F(X) \leq z\} = \mathrm{P}\left\{X \leq F^{-1}(z)\right\} = F\left\{F^{-1}(z)\right\} = z = F(z)$$

and since $F'(z) = f(z)$ we finally have that $Z = F(X) \sim U[0,1]$. The proof of the theorem will follow from this and the relation

$$\Lambda_{\mathrm{X}}(x) = -\log(1 - F(x)).$$

For $y < 0$, $F(y) = 0$ and therefore $\Lambda_{\mathrm{X}}(y) = 0$. For $y \geq 0$ we have

$$\begin{aligned}\mathrm{P}\{Y \leq y\} &= \mathrm{P}\{\Lambda_{\mathrm{X}}(X) \leq y\} = \mathrm{P}\{-\log(1 - F(X)) \leq y\} = \mathrm{P}\{\log(1 - F(X)) \geq -y\} \\ &= \mathrm{P}\left\{1 - F(X) \geq e^{-y}\right\} = \mathrm{P}\left\{F(X) \leq 1 - e^{-y}\right\} = 1 - e^{-y}\end{aligned}$$

since $F(x)$ is distributed as $U[0,1]$. Using again the relation $F'(y) = f(y)$ we get that $Y = \Lambda_{\mathrm{X}}(X) \sim \exp(1)$. ∎

Theorem 4.2.2. *Let X be a continuous random variable. Suppose that the transformation $y = g(x)$, $x \in R_x$, $R_x \subseteq \mathbb{R}$, $y \in R_y = g(R_x)$ is 1-1 from R_x to R_y. Suppose also that $\frac{d}{dy}g^{-1}(y)$ exists and is continuous and let $\lambda_Y(y)$ denote the hazard rate of Y. Then the random variable $Y = g(X)$ is continuous with hazard rate*

$$\lambda_Y(y) = \lambda_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y).$$

Practically this means that a sample from any pdf f can be transformed to a sample from the standard exponential distribution by using as transformation the integrated hazard rate of the density f. As in our case the function Λ_{X} is unknown, we use as the transformation a smooth estimate of the integrated hazard rate such as,

$$\hat{\Lambda}_{1,\mathrm{X}}(x) = \int_0^x \hat{\lambda}_{T,1}(u|h_1)\, du.$$

Taking $g = \hat{\Lambda}_{1,\mathrm{X}}$, we see that the hazard rate of the transformed (target) sample, $\lambda_Y(y;g)$, can be estimated accurately by an ordinary kernel hazard rate estimator such as

$$\hat{\lambda}_Y(y;g,h_2) = \sum_{i=1}^{n} \frac{K_{h_2}(y - Y_{(i)})}{n-i+1}.$$

Then, by change of variable, we can back transform $\hat{\lambda}_Y$ to an estimator of the original hazard rate,

$$\begin{aligned}\hat{\lambda}_{T,2}(x|h_1,h_2) \equiv \hat{\lambda}_X(x;g;h_1,h_2) &= \hat{\lambda}_Y\left(g(x);g,h_2\right) g'(x) \\ &= \sum_{i=1}^{n} \frac{K_{h_2}\left(g(x) - g(X_{(i)})\right)}{n-i+1} g'(x),\end{aligned}$$

which we call as the second step transformed kernel hazard rate estimator (TKHRE). Now, if g were exactly Λ_{X} so that λ_Y were exactly standard exponential hazard rate, $\hat{\lambda}_Y$ would be unbiased for y from $g(h)$ and on and $\hat{\lambda}_X$ would be unbiased for x from $g^{-1}(h)$ and on. Since we use just an estimate of the integrated hazard rate as the transformation, the bias is not completely eliminated. However, in the next section we show that with the current setup and with appropriate choice of bandwidths h_1 and h_2, the square error of $\hat{\lambda}_X(x;\hat{\Lambda}_{1,\mathrm{X}}(x;h_1);h_2)$ is of order $O_p(n^{-8/9})$ rather than $O_p(n^{-4/5})$ as for an ordinary kernel hazard rate estimator. We note here that more bias reduction is possible by iterating the process, with the transformation in the next step being the indefinite integral of $\hat{\lambda}_{T,2}(x|h_1,h_2)$ and so on.

The above proposal has been implemented in density settings by Ruppert and Cline [35]. There, the uniform distribution plays the role played here by the exponential distribution. This work is an extension of their idea to the hazard rate case.

4.3 Asymptotic error of the estimator.

Before we give the proof we note that in order to keep notation as simple as possible we suppress the dependence of $\hat{\lambda}_{T,1}$ on the bandwidth, i.e. we write $\hat{\lambda}_{T,1}(\cdot)$ instead of $\hat{\lambda}_{T,1}(\cdot|h)$. For the same reason we simply write $\hat{\Lambda}_1(\cdot)$ instead of $\hat{\Lambda}_{1,\mathrm{X}}(\cdot)$. Now, let

$$T = \sup\left\{x|1 - F_X(x) > \varepsilon\right\}, \quad \varepsilon > 0.$$

Also let $\Delta = h(\log n)^{\frac{1}{2}}$, $C > 0$ a constant and x' to be a point in the interior of the support of X. Define the set $\mathcal{G}$ to be

$$\begin{aligned}\mathcal{G} = \Big\{g : g \text{ is a fifth degree polynomial}, g(x) = 0, &\left|g^{(m)}(x) - \Lambda_{\mathrm{X}}^{(m)}(x)\right| \leq \Delta \\ \text{for } m = 1,2,3,4 \text{ and } &\left|g^{(5)}(x) - \Lambda_{\mathrm{X}}^{(5)}(x)\right| \leq C\Big\}.\end{aligned}$$

Denote with F_n the empirical distribution function of the original sample which was defined in the introduction. Define also

$$\tilde{\Lambda}_1(x') = \sum_{m=1}^{5} \hat{\Lambda}_1^{(m)}(x) \frac{(x'-x)^m}{m!},$$

i.e. $\tilde{\Lambda}_1$ is an approximation of $\hat{\Lambda}_1$ based on a fifth order Taylor expansion of $\hat{\Lambda}_1$ around x'. Note that the summation in the above definition starts at $m = 1$ and not at $m = 0$ so that $\tilde{\Lambda}_1$ is in $\mathcal{G}$.

We will prove that the second step TKHRE is an estimator with the same rate of convergence as of that we get by using a fourth order kernel estimate. The following assumptions are used throughout this section.

B.1 $\lambda_X(x)$ is bounded and has four bounded derivatives in a neighborhood of x and Λ_X is five time continuously differentiable on $[0, T]$.

B.2 There exists $\varepsilon > 0$ such that for every x in $[\,0, T\,]$, $F_X(x) < 1 - \varepsilon$.

B.3 K is a Lipschitz continuous (i.e. there is a positive real constant L such that $|K(x) - K(y)| \leq L|x - y|$ for every x and y in the support of the kernel) symmetric kernel with support $[\,-1, 1]$, and has five continuous derivatives.

B.4 $\sup |F_n(\hat{\Lambda}_1(X_i)) - F_n(\tilde{\Lambda}_1(X_i))| = O_p\left((\sqrt{n})^{-1}\right)$ for every $i = 1, \ldots, n$.

Then we have the following theorem

Theorem 4.3.1. *Suppose that $G(x) = -\log(1 - x)$ so that $\Lambda_{\mathrm{X}} = G \circ F_X$ on $[\,0, T\,]$ and let $h_0 = n^{-\frac{1}{9}}$. Suppose also that for $i = 1, 2$,*

$$h_i/h_0 \to c_i > 0 \text{ as } n \to +\infty.$$

Then for each $M > 0$ we have

$$\sup_{|x-x'| \leq Mh_0} \left|\hat{\lambda}_{T,2}(x'|h_1, h_2) - \lambda_X(x')\right| = O_p\left(h_2^4\right).$$

Remark 4.1. The technicalities involved in the proof are similar to those in the density estimation case, given in Ruppert and Cline [35]. We establish the convergence of $\hat{\lambda}_{T,2}$ at each x over a shrinking neighborhood of x because in order to study the second step TKHRE we need to ensure that $\hat{\lambda}_{T,1}$ converges uniformly in a neighborhood of x.

Proof. Since $G(F_X(x)) = -\log(1 - F_X(x))$ has a positive derivative at x and since $\Delta \to 0$ there exists $N > 0$ and $M_1 > 0$ such that for all $n > N$, for all x' and x'' in $[\,x - Mh, x + Mh\,]$ and for all g in $\mathcal{G}$ we have,

$$|g(x') - g(x'')| \geq \frac{1}{M_1}|x' - x''|. \tag{4.2}$$

Throughout the proof we will use the constants A_1 and A_2 which satisfy $A_1 = A_2 + c_2'M_1$, $c_2' = 2c_2$. Note that $h_i \leq c_2'h_0$ for n large. We will prove the theorem with fixed $M = A_1$. From the theorem of transformation of random variables we have

$$\hat{\lambda}_X(x';g;h_1,h_2) = \hat{\lambda}_Y(g(x');g,h_2)g'(x') \tag{4.3}$$
$$\text{and} \;\; \lambda_X(x') = \lambda_Y(g(x');g)g'(x'). \tag{4.4}$$

For $g = \hat{\Lambda}_1$ we have,

$$\sup_{|x-x'|\le A_2h_0} \left| \hat{\lambda}_{T,2}(x'|h_1,h_2) - \lambda_X(x') \right| = \sup_{|x-x'|\le A_2h_0} \left| \hat{\lambda}_X\left(x';\hat{\Lambda}_1;h_1,h_2\right) - \lambda_X(x') \right| =$$
$$\sup_{|x-x'|\le A_2h_0} \left| \left\{ \hat{\lambda}_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1,h_2\right) - \lambda_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1\right) \right\} \hat{\Lambda}_1'(x') \right| \le$$
$$\sup_{|x-x'|\le A_2h_0} \left| \hat{\lambda}_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1,h_2\right) - \lambda_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1\right) \right| \left| \hat{\Lambda}_1'(x') \right|.$$

Notice that $\hat{\Lambda}_1' = \hat{\lambda}_{T,1}$ is bounded because both the kernel and the denominator are bounded. Therefore we only need to show that

$$\sup_{|x-x'|\le A_2h_0} \left| \hat{\lambda}_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1,h_2\right) - \lambda_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1\right) \right| = O_p\left(h_2^4\right).$$

Writing

$$\sup_{|x-x'|\le A_2h_0} \left| \hat{\lambda}_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1,h_2\right) - \lambda_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1\right) \right| \le$$
$$\sup_{|x-x'|\le A_2h_0} \left| \hat{\lambda}_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1,h_2\right) - \hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right) \right| +$$
$$\sup_{|x-x'|\le A_2h_0} \left| \hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right) - \lambda_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1\right) \right| +$$
$$\sup_{|x-x'|\le A_2h_0} \left| \lambda_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1\right) - \lambda_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1\right) \right|.$$

we see that the theorem will be proved if we show that

(a)
$$\sup_{|x-x'|\le A_2h_0} \left| \hat{\lambda}_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1,h_2\right) - \hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right) \right| = O_p\left(h_2^4\right)$$

(b)
$$\sup_{|x-x'|\le A_2h_0} \left| \hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right) - \lambda_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1\right) \right| = O_p\left(h_2^4\right)$$

(c)
$$\sup_{|x-x'|\le A_2h_0} \left| \lambda_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1\right) - \lambda_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1\right) \right| = O_p\left(h_2^4\right).$$

We start with (a). First, note that from lemma 4.3.4 on page 82 we have

$$\sup_{|x-x'|\le A_1h_0} \left| \hat{\Lambda}_1(x') - \tilde{\Lambda}_1(x') \right| \le \sup_{|x-x'|\le A_1h_0} \left| \hat{\Lambda}_1^{(5)}(x') \right| (A_1h_0)^5 = O_p(h_2^5).$$

Now,

$$\sup_{|x-x'|\leq A_2h_0}\left|\hat{\lambda}_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1,h_2\right)-\hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right)\right|\leq$$

$$\sup_{|x'-x|\leq A_1h_0}\frac{1}{nh_2}\sum_{i=1}^{n}\left|\frac{K\left(\frac{\hat{\Lambda}_1(X_i)-\hat{\Lambda}_1(x')}{h_2}\right)}{1-F_n(\hat{\Lambda}_1(X_i))}-\frac{K\left(\frac{\tilde{\Lambda}_1(X_i)-\tilde{\Lambda}_1(x')}{h_2}\right)}{1-F_n(\tilde{\Lambda}_1(X_i))}\right|I_{\{|X_i-x'|\leq M_1h_2\}}.$$

Using Lipschitz continuity of the kernel,

$$\sup_{|x'-x|\leq A_1h_0}\left|K\left(\frac{\hat{\Lambda}_1(X_i)-\hat{\Lambda}_1(x')}{h_2}\right)-K\left(\frac{\tilde{\Lambda}_1(X_i)-\tilde{\Lambda}_1(x')}{h_2}\right)\right|\leq$$

$$\sup_{|x'-x|\leq A_1h_0}\frac{2}{h_2}\left|\hat{\Lambda}_1(x')-\tilde{\Lambda}_1(x')\right|.$$

By B.4 and by working as in section 3.6, we have that

$$\frac{K\left(\frac{\hat{\Lambda}_1(X_i)-\hat{\Lambda}_1(x')}{h_2}\right)}{1-F_n(\hat{\Lambda}_1(X_i))}-\frac{K\left(\frac{\tilde{\Lambda}_1(X_i)-\tilde{\Lambda}_1(x')}{h_2}\right)}{1-F_n(\tilde{\Lambda}_1(X_i))}=$$

$$\frac{K\left(\frac{\hat{\Lambda}_1(X_i)-\hat{\Lambda}_1(x')}{h_2}\right)}{1-F(\hat{\Lambda}_1(X_i))}\left\{\frac{1}{1-F_n(\tilde{\Lambda}_1(X_i))}-\frac{1}{1-F(\tilde{\Lambda}_1(X_i))}\right\}-$$

$$\frac{K\left(\frac{\tilde{\Lambda}_1(X_i)-\tilde{\Lambda}_1(x')}{h_2}\right)}{1-F(\tilde{\Lambda}_1(X_i))}\left\{\frac{1}{1-F_n(\tilde{\Lambda}_1(X_i))}-\frac{1}{1-F(\tilde{\Lambda}_1(X_i))}\right\}=$$

$$\frac{K\left(\frac{\hat{\Lambda}_1(X_i)-\hat{\Lambda}_1(x')}{h_2}\right)}{1-F(\hat{\Lambda}_1(X_i))}-\frac{K\left(\frac{\tilde{\Lambda}_1(X_i)-\tilde{\Lambda}_1(x')}{h_2}\right)}{1-F(\tilde{\Lambda}_1(X_i))}+o_p\left(\frac{1}{\sqrt{n}}\right).$$

Thus, for some positive generic constant C_1,

$$\sup_{|x-x'|\leq A_2h_0}\left|\hat{\lambda}_Y\left(\hat{\Lambda}_1(x');\hat{\Lambda}_1,h_2\right)-\hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right)\right|\leq$$

$$C_1\sup_{|x'-x|\leq A_1h_0}\left|\hat{\Lambda}_1(x')-\tilde{\Lambda}_1(x')\right|\sup_{|x'-x|\leq A_1h_0}\frac{1}{nh_2^2}\sum_{i=1}^{n}I_{\{|X_i-x'|\leq M_1h_2\}}=O_p(h_2^4).$$

To prove (b), first write,

$$\sup_{|x-x'|\leq A_2h_0}\left|\hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right)-\lambda_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1\right)\right|=$$

$$\sup_{|x-x'|\leq A_2h_0}\left|\hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right)-\mathbb{E}\hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right)\right|+$$

$$\sup_{|x-x'|\leq A_2h_0}\left|\mathbb{E}\hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right)-\lambda_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1\right)\right|.$$

We will prove that the RHS of the above inequality is of order $O_p(h_2^4)$. The proof of

$$\sup_{|x-x'|\leq A_2h_0}\left|\hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right)-\mathbb{E}\hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right)\right|=O_p(h_2^4)$$

is direct application of lemma 4.3.1 on page 76 with $g=\tilde{\Lambda}_1$ and $h=h_2$. To prove that

$$\sup_{|x-x'|\leq A_2h_0}\left|\mathbb{E}\hat{\lambda}_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1,h_2\right)-\lambda_Y\left(\tilde{\Lambda}_1(x');\tilde{\Lambda}_1\right)\right|=O_p(h_2^4)$$

first, denote with $F_{Y,n}$ the empirical cdf of the sample Y_i, $i=1,\ldots,n$,

$$F_{Y,n}(y)=\frac{\{\#Y_i\leq y\}-1}{n}.$$

Then for g in $\mathcal{G}$,

$$\begin{aligned}\mathbb{E}\hat{\lambda}_Y\left(g(x');g,h_2\right)&=\mathbb{E}\frac{1}{h_2}\sum_{i=1}^{n}\frac{K\left(\frac{g(x')-g(X_i)}{h_2}\right)}{1-F_{Y,n}(g(X_i))}=\mathbb{E}\frac{1}{h_2}\sum_{i=1}^{n}\frac{K\left(\frac{g(x')-Y_i}{h_2}\right)}{1-F_{Y,n}(Y_i)}\\&\simeq\frac{1}{h_2}\int K\left(\frac{g(x')-u}{h_2}\right)\lambda_Y(u)\,du=\int K(z)\lambda_Y\left(g(x')-h_2z;g\right)\,dz\end{aligned}\tag{4.5}$$

after applying the change of variable $g(x')-u=h_2z$. Using the fact that for any function f the remainder R_n of a Taylor expansion of $f(x+h)$ around h up to the $n-1$ power is

$$R_n=\frac{h^n}{n!}f^{(n)}(x+\theta h),\quad\theta<1$$

and expanding $\lambda_Y\left(g(x')-h_2z;g\right)$ in Taylor series around h_2z up to the first term yields

$$\mathbb{E}\hat{\lambda}_Y\left(g(x');g,h_2\right)-\lambda_Y\left(g(x');g\right)=\frac{h_2^2}{2}\int z^2K(z)\lambda_Y''\left(g(x')+h_2\theta z\right)\,dz.\tag{4.6}$$

If we take $|x-x'|\leq A_2h_0$ and $|x-x''|\leq(A_2+c_2'M_1)h_0=A_1h_0$, with M_1 being as in (4.2), then from the mean value theorem we have that $g(x')+h_2\theta z=g(x'')$. A bound for λ_Y'' in (4.6) can then be obtained by using the fact that the hazard rate of the exponential distribution is constant and therefore all its derivatives are zero, i.e.

$$\lambda_Y^{(l)}\left(\Lambda_{\mathrm{X}}\left(x'\right);\Lambda_{\mathrm{X}}\right)=0\ \text{ for }\ l>0.\tag{4.7}$$

In lemma 4.3.3, page 79 we prove that for $g=\tilde{\Lambda}_1$,

$$\sup_{|x-x'|\leq A_2h_0}\left|\lambda_Y''\left(g(x');g\right)\right|=\sup_{|x-x'|\leq A_2h_0}\left|\lambda_Y''\left(g(x');g\right)-\lambda_Y''\left(\Lambda_{\mathrm{X}}(x');\Lambda_{\mathrm{X}}\right)\right|=O_p\left(h_2^2\right).\tag{4.8}$$

Substituting (4.7) with $l=2$ in (4.8) gives

$$\sup_{|x-x'|\leq A_2h_0}\left|\lambda_Y''\left(g(x');g\right)\right|=O_p\left(h_2^2\right).\tag{4.9}$$

Applying (4.9) to λ_Y'' in (4.6) yields

$$\left|\mathbb{E}\hat{\lambda}_Y\left(g(x');g,h_2\right)-\lambda_Y\left(g(x');g\right)\right|=O_p\left(h_2^2h_2^2\right)=O_p\left(h_2^4\right).\tag{4.10}$$

Obviously, (4.10) with $g = \tilde{\Lambda}_1$ proves (b). The proof of (c) is very similar to the proof of (5.22) in [35]. In our case, by (4.3) it reduces to

$$\sup_{|x-x'|\le A_2 h_0} \left| \lambda_Y\left(\hat{\Lambda}_1(x'); \hat{\Lambda}_1\right) - \lambda_Y\left(\tilde{\Lambda}_1(x'); \tilde{\Lambda}_1\right) \right| =$$

$$\sup_{|x-x'|\le A_2 h_0} \left| \frac{\lambda_X(x')}{\hat{\Lambda}_1'(x')} - \frac{\lambda_X(x')}{\tilde{\Lambda}_1'(x')} \right| = \sup_{|x-x'|\le A_2 h_0} \left| \frac{\lambda_X(x')}{\hat{\Lambda}_1'(x')\tilde{\Lambda}_1'(x')} \left\{ \hat{\Lambda}_1'(x') - \tilde{\Lambda}_1'(x') \right\} \right|.$$

From B.1 we have that λ_X is bounded. Recall from page 73 that $\hat{\Lambda}_1'(x')$ is also bounded. From the definition of $\tilde{\Lambda}_1'(x')$ it is easy to see that it is bounded as a finite sum of bounded terms. Therefore there is an appropriate constant C such that

$$\sup_{|x-x'|\le A_2 h_0} \left| \lambda_Y\left(\hat{\Lambda}_1(x'); \hat{\Lambda}_1\right) - \lambda_Y\left(\tilde{\Lambda}_1(x'); \tilde{\Lambda}_1\right) \right| \le C \sup_{|x-x'|\le A_2 h_0} \left| \hat{\Lambda}_1'(x') - \tilde{\Lambda}_1'(x') \right| = O_p(h_2^4)$$

by lemma 4.3.4 on page 82. ∎

4.3.1 Lemmas.

Lemma 4.3.1. *Suppose that $h = cn^{-\frac{1}{9}}$ for some $c > 0$ and fix $M > 0$. Then,*

$$\sup_{|x'-x|\le Mh} \sup_{g\in\mathcal{G}} \left| \hat{\lambda}_Y\left(g(x'); g, h\right) - \mathbb{E}\hat{\lambda}_Y\left(g(x'); g, h\right) \right| = O_p\left(h^4\right).$$

Proof. The lemma is essentially an application of theorem 37 of Pollard [32]. We first show that the theorem is applicable in this case. In Pollard's notation let the probability measure, Q, be the usual empirical measure

$$\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(X_i)$$

and the permissible class of functions, $\mathcal{F}_n$, be

$$\mathcal{D} = \{\kappa(\cdot; x, g) : g \in \mathcal{G} \text{ and } x \in [x - Mh, x + Mh]\}$$

with

$$\kappa(\cdot; x, g) = \frac{K\left(\frac{g(\cdot)-g(x')}{h}\right) - K\left(\frac{\Lambda_X(\cdot)-\Lambda_X(x')}{h}\right)}{1 - F_n(g(\cdot))}$$

where F_n is the empirical estimate of the cdf F_X defined in the introduction. Since

$$\frac{K\left(\frac{g(\cdot)-g(x')}{h}\right) - K\left(\frac{\Lambda_X(\cdot)-\Lambda_X(x')}{h}\right)}{1 - F_n(g(\cdot))} = \frac{K\left(\frac{g(\cdot)-g(x')}{h}\right) - K\left(\frac{\Lambda_X(\cdot)-\Lambda_X(x')}{h}\right)}{1 - F_X(g(\cdot))} + o_p\left(\frac{1}{\sqrt{n}}\right),$$

in order to avoid a longer technical argument, we will prove the lemma with

$$\kappa(\cdot; x, g) = \frac{K\left(\frac{g(\cdot)-g(x')}{h}\right) - K\left(\frac{\Lambda_X(\cdot)-\Lambda_X(x')}{h}\right)}{1 - F_X(g(\cdot))}.$$

Also, fix $\eta > 0$ and take

$$\{a_n\} = \frac{(\log n)^{1+\eta}}{\sqrt{n\delta_n^2}}$$

to be the non-increasing sequence of positive numbers with $\delta_n = \sqrt{M_7}\sqrt{h}\,\Delta$, where M_7 is a positive generic constant. We immediately see that

$$\frac{n\delta_n^2 a_n^2}{\log n} \to +\infty, \quad \text{as } n \to +\infty$$

as required by the conditions of the theorem. We also have to show that

(i)

$$\mathbb{E}\kappa^2(X_1; x', g) \leq \delta_n^2 \;\text{ for each } \kappa \text{ in } \mathcal{D}$$

(ii)

$$\sup_Q N_1(\varepsilon, \mathcal{D}) \leq A\varepsilon^{-6}, \quad 0 < \varepsilon < 1$$

where N_1 is the set of covering numbers of D (see Pollard [32], pp. 25 for a definition) and A is a constant independent of n. Starting with (i) we have,

$$\begin{aligned}\mathbb{E}\kappa^2(X_1; x', g) &= \int \left\{ \frac{K\left(\frac{g(u)-g(x')}{h}\right) - K\left(\frac{\Lambda_X(u)-\Lambda_X(x')}{h}\right)}{1 - F_X(g(u))} \right\}^2 f_X(g(u))\, dg(u) \\ &= \int \left\{ K\left(\frac{g(u)-g(x')}{h}\right) - K\left(\frac{\Lambda_X(u)-\Lambda_X(x')}{h}\right) \right\}^2 \\ &\quad \times \frac{\lambda_X(g(u))}{1 - F_X(g(u))}\, dg(u) \end{aligned} \tag{4.11}$$

Observe that for

$$|x' - x''| \leq M_1 h, \;\; |x - x'| \leq (M + M_1)h \;\text{ and }\; |x'' - x| \leq (M + M_1)h,$$

we have

$$\begin{aligned}|g(x'') - g(x') - \Lambda_X(x'') + \Lambda_X(x')| &= |(g(x'') - \Lambda_X(x'')) - (g(x') - \Lambda_X(x'))| \\ &\leq |g(x'') - \Lambda_X(x'')| - |g(x') - \Lambda_X(x')| \leq \int_{x'}^{x''} |g'(u) - \Lambda_X'(u)|\, du.\end{aligned}$$

From the mean value theorem there is ξ in $[\,x', x''\,]$ and a constant $M_4 > 0$ such that

$$\int_{x'}^{x''} |g'(u) - \Lambda_X'(u)|\, du = |x'' - x'|\, |g'(\xi) - \Lambda_X'(\xi)| \leq M_4 h\Delta.$$

Now, since the kernel is a Lipschitz continuous function,

$$\left| K\left(\frac{g(x'')-g(x')}{h}\right) - K\left(\frac{\Lambda_X(x'')-\Lambda_X(x')}{h}\right) \right| \leq \frac{1}{h}\int_{x'}^{x''} |g'(u) - \Lambda_X'(u)|\, du \leq \frac{1}{h} M_5 h\Delta \tag{4.12}$$

with M_5 being positive generic constant. Also, from B.1 and B.2 we have that

$$\frac{\lambda_X(g(u))}{1-F_X(g(u))}$$

is bounded. This, together with (4.11) and (4.12) implies that for a suitable positive constants M_6 and for $|X_1 - x'| \leq M_1 h$,

$$\mathbb{E}\kappa^2(X_1;x',g) \leq \frac{1}{h^2} M_6^3 h^2 \Delta^2 h \int \frac{\lambda_X(g(u))}{1-F_X(g(u))}\, dg(u) \leq M_7 h \Delta^2 = \delta_n^2.$$

This completes the proof of (i). For the proof of (ii) define the set

$$\mathcal{Y}(\varepsilon) = [\, x - Mh, x + Mh \,] \cap (h \in j) : j \in \mathbb{Z}$$

and let $\mathcal{G}_\varepsilon$ be the set of all g^* in $\mathcal{G}$ with coefficients of the form

$$a_m^* = \Lambda_X^{(m)}(x) + j\Delta\varepsilon,\ m = 1,2,3,4 \ \text{ and } \ a_5^* = \Lambda_X^{(5)}(x) + jC\varepsilon, |j| \leq [\,\varepsilon^{-1}\,] + 1,\ j \in \mathbb{Z}.$$

Then (ii) will be proved, as in [35] by covering $\mathcal{D}$ by $L_1(\Lambda_{\mathrm{X}})$ balls with centers in the set

$$\mathcal{D}_\varepsilon = \{\kappa\,(\cdot; x', g) : g \in \mathcal{G}_\varepsilon \text{ and } x' \in \mathcal{Y}(\varepsilon)\}$$

and radii equal to $M_8\varepsilon$ where M_8 is a positive generic constant. The proof then reduces to proving that, for positive generic constants M_9 and M_{10},

$$\mathrm{card}[\,\mathcal{G}_\varepsilon\,] \leq M_9 \varepsilon^{-5} \ \text{ and } \ \mathrm{card}[\,\mathcal{Y}(\varepsilon)\,] \leq M_{10}\varepsilon^{-1}.$$

In both cases the proof is analogous to that of [35] and therefore is omitted. Now, set

$$\hat\lambda_{Y,*}(y;g,h) = \hat\lambda_Y(y;g,h) - \mathbb{E}\hat\lambda_Y(y;g,h).$$

Then, applying Pollard's theorem,

$$\sup_{|x'-x|\leq Mh} \sup_{g \in \mathcal{G}} \left| \hat\lambda_{Y,*}(g(x');g,h) - \hat\lambda_{Y,*}(\Lambda_{\mathrm{X}}(x');g,h) \right| =$$

$$\sup_{|x-x'|\leq Mh} \sup_{g \in \mathcal{G}} \left| \frac{1}{h} \sum_{i=1}^{n} \kappa\,(X_i;x',g) - \mathbb{E}\kappa\,(X_i;x',g) \right| = o\left(\frac{1}{h}\delta_n^2 a_n\right) = o\left(h^4\right) \tag{4.13}$$

almost surely as $n \to \infty$. From lemma 4.3.2, with $h = 1/n^{-\frac{1}{9}}$

$$\sup_{|x-x'|\leq Mh} \sup_{g \in \mathcal{G}} \left| \hat\lambda_{Y,*}(\Lambda_{\mathrm{X}}(x');g,h) \right| = O_p\left(\frac{1}{\sqrt{nh}}\right) = O_p\left(h^4\right) \tag{4.14}$$

and therefore, from (4.13) and (4.14) we conclude that

$$\sup_{|x-x'|\leq Mh} \sup_{g \in \mathcal{G}} \left| \hat\lambda_{Y,*}(g(x');g,h) \right| = O_p\left(h^4\right). \quad \blacksquare$$

Lemma 4.3.2. *Let $s \in [0,1]$ and suppose that $h = h_n$ is a sequence such that $h > 0$, $h \to 0$ and $nh \to +\infty$. Fix $D > 0$. Define the process*

$$X_n(s) = \sqrt{nh}\left\{\hat{\lambda}_{T,1}\left(x + Dh\left(s - \frac{1}{2}\right)\right) - \mathbb{E}\hat{\lambda}_{T,1}\left(x + Dh\left(s - \frac{1}{2}\right)\right)\right\}$$

and let X be a stationary zero-mean Gaussian process X on $[0,1]$ with

$$\mathbb{C}\text{ov}\left(X(s), X(s')\right) = \frac{\lambda_X(x)}{1 - F_X(x)}\int K(u + D(s - s'))K(u)\,du.$$

Then X_n converges weakly to X in $C[0,1]$ as $n \to +\infty$ (i.e. if F_n and F are the distribution functions of X_n and X respectively then $\lim_{n\to+\infty} F_n(t) = F(t)$ for every continuity point t of F) and

$$\sup_{s\in[0,1]} |X_n(s)| = O_p(1) \ \text{ as } n \to \infty.$$

Proof. The proof is similar to the proof of lemma 5.1 in Ruppert and Cline [35] for $l = 0$ with the only difference being that (in Ruppert and Cline's notation) in order to show that X_n is tight we should use a stronger inequality to bound $\mathbb{E}(X_n(s) - X_n(s'))^2$. According to a version of Rosenthal's inequality for sums of independent random variables with $\mathbb{E}X_i = 0$, $i = 1, \ldots, n$

$$\mathbb{E}\left|\sum_{i=1}^{n} X_i\right|^p \le c(p)\left\{\sum_{i=1}^{n}\mathbb{E}|X_i|^p + \left(\sum_{i=1}^{n}\mathbb{E}X_i^2\right)^{\frac{p}{2}}\right\}$$

for some constant c which depends on p. In the case that $p = 2$ the inequality becomes

$$\mathbb{E}\left|\sum_{i=1}^{n} X_i\right|^2 \le 2c\sum_{i=1}^{n}\mathbb{E}|X_i|^2.$$

Now,

$$\mathbb{E}(X_n(s) - X_n(s'))^2 = nh\mathbb{E}\left\{\hat{\lambda}_{T,1}\left(x + Dh\left(s - \frac{1}{2}\right)\right) - \hat{\lambda}_{T,1}\left(x + Dh\left(s' - \frac{1}{2}\right)\right)\right\}^2$$

which from the definition of $\hat{\lambda}_{T,1}$ can be written as

$$\mathbb{E}(X_n(s) - X_n(s'))^2 = \frac{1}{nh}\mathbb{E}\left\{\sum_{i=1}^{n}\left(\frac{K\left(\frac{x-X_i}{h} - D\left(s - \frac{1}{2}\right)\right)}{1 - F_n(X_i)} - \frac{K\left(\frac{x-X_i}{h} - D\left(s' - \frac{1}{2}\right)\right)}{1 - F_n(X_i)}\right)\right\}^2.$$

Define $Y^*_{n,i}(s)$ as

$$Y^\star_{n,i}(s) = \frac{1}{\sqrt{nh}}\frac{K\left(\frac{x-X_i}{h} - D\left(s + \frac{1}{2}\right)\right)}{1 - F_n(X_i)}.$$

Now,

$$\frac{K\left(\frac{x-X_i}{h} - D\left(s + \frac{1}{2}\right)\right)}{1 - F_n(X_i)} = \frac{K\left(\frac{x-X_i}{h} - D\left(s + \frac{1}{2}\right)\right)}{1 - F_X(X_i)} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

and so, it is asymptotically equivalent to use

$$Y_{n,i}(s) = \frac{1}{\sqrt{nh}} \frac{K\left(\frac{x-X_i}{h} - D\left(s+\frac{1}{2}\right)\right)}{1-F_X(X_i)}.$$

Thus we can write

$$\mathbb{E}(X_n(s) - X_n(s'))^2 = \mathbb{E}\left(\sum_{i=1}^{n} \{Y_{n,i}(s) - Y_{n,i}(s')\}\right)^2.$$

Then, from Rosenthal's inequality

$$\mathbb{E}(X_n(s) - X_n(s'))^2 \leq c \sum_{i=1}^{n} \mathbb{E}\left(Y_{n,i}(s) - Y_{n,i}(s')\right)^2.$$

Now,

$$\sum_{i=1}^{n} \mathbb{E}\left(Y_{n,i}(s) - Y_{n,i}(s')\right)^2$$

$$= \frac{1}{h} \int \left\{ \frac{K\left(\frac{x-y}{h} + D\left(s-\frac{1}{2}\right)\right) - K\left(\frac{x-y}{h} + D\left(s'-\frac{1}{2}\right)\right)}{1-F_X(X_i)} \right\}^2 f_X(y)\, dy =$$

$$= \frac{1}{h} \int \left\{ K\left(\frac{x-y}{h} + D\left(s-\frac{1}{2}\right)\right) - K\left(\frac{x-y}{h} + D\left(s'-\frac{1}{2}\right)\right) \right\}^2 \frac{\lambda_X(y)}{1-F_X(y)}\, dy.$$

Applying the change of variable

$$\frac{x-y}{h} + D(s'-\frac{1}{2}) = u$$

the above equation becomes

$$\mathbb{E}\sum_{i=1}^{n}(Y_{n,i}(s) - Y_{n,i}(s'))^2 = \int (K\left(u + D\left(s-s'\right)\right) - K(u))^2 \frac{\lambda_X\left(x - hu - Dh\left(s'-\frac{1}{2}\right)\right)}{1 - F_X\left(x - hu - Dh\left(s'-\frac{1}{2}\right)\right)}\, du.$$

From Taylor's theorem, for $\theta < 1$, $K\left(u + D\left(s-s'\right)\right)$ can be written as

$$K\left(u + D\left(s-s'\right)\right) = K(u) + D(s-s')K'(u + \theta D(s-s'))$$

and so

$$\mathbb{E}\sum_{i=1}^{n}\left(Y_{n,i}(s) - Y_{n,i}(s')\right)^2 = (D(s-s'))^2 \int \left(K'(u+\theta D(s-s'))\right)^2 \times$$

$$\frac{\lambda_X\left(x - hu - Dh\left(s'-\frac{1}{2}\right)\right)}{1 - F_X\left(x - hu - Dh\left(s'-\frac{1}{2}\right)\right)}\, du \leq M(s-s')^2$$

for some positive M, from which it follows that

$$\mathbb{E}\left(X_n(s) - X_n(s')\right)^2 \leq M(s-s')^2$$

uniformly in n, s and s'. ∎

Lemma 4.3.3. *Suppose that $g = \tilde{\Lambda}_1$. Then, under the conditions of theorem 4.3.1*

$$\sup_{|x'-x|<A_2h_0} \left|\lambda_Y''\left(g(x');g\right) - \lambda_Y''\left(\Lambda_{\mathrm{X}}(x');\Lambda_{\mathrm{X}}\right)\right| = O_p\left(h_2^2\right).$$

Proof. Differentiating (4.4) twice and solving for λ_Y'', gives

$$\lambda_Y''\left(g\left(x'\right);g\right) = \frac{1}{(g'(x'))^3}\sum_{m=0}^{2} Q_m(x';g)\lambda_X^{(m)}(x')$$

where the Q_m's, $m = 0,1,2$ are functions of $g^{(0)} = 1, g'$ and g''. Similarly, $\lambda_Y''\left(\Lambda_{\mathrm{X}}(x');\Lambda_{\mathrm{X}}\right)$ can be written in the same fashion with g replaced by Λ_{X}. Then,

$$\sup_{|x-x'|\le A_2h_0} \left|\lambda_Y''\left(g(x');g\right) - \lambda_Y''\left(\Lambda_{\mathrm{X}}(x');\Lambda_{\mathrm{X}}\right)\right| =$$

$$\sup_{|x-x'|\le A_2h_0}\left|\frac{1}{g'(x')^3}\sum_{m=0}^{2} Q_m(x',g)\lambda_X^{(m)}(x') - \frac{1}{\lambda_X^3(x')}\sum_{m=0}^{2} Q_m(x',\Lambda_{\mathrm{X}})\lambda_X^{(m)}(x')\right|$$

$$\le \sup_{|x-x'|\le A_2h_0}\sum_{m=0}^{2}\left|\lambda_X^{(m)}(x')\right|\left|\frac{Q_m(x',g)}{g'(x')^3} - \frac{Q_m(x',\Lambda_{\mathrm{X}})}{\lambda_X^3(x')}\right|.$$

From assumption B.1, $|\lambda_X^{(m)}(x')|$ is bounded for $m = 0,1,2$, therefore there exist constant M such that

$$\sup_{|x-x'|\le A_2h_0} \left|\lambda_Y''\left(g(x');g\right) - \lambda_Y''\left(\Lambda_{\mathrm{X}}(x');\Lambda_{\mathrm{X}}\right)\right| \le M \sup_{|x-x'|\le A_2h_0}\sum_{m=1}^{3}\left|g^{(m)}(x') - \Lambda_{\mathrm{X}}^{(m)}(x')\right|.$$

Now, for $g \equiv \tilde{\Lambda}_1$,

$$\sup_{|x-x'|\le A_2h_0}\sum_{m=1}^{3}\left|g^{(m)}(x') - \Lambda_{\mathrm{X}}^{(m)}(x')\right| \le \sup_{|x-x'|\le A_2h_0}\left|\tilde{\Lambda}_1'(x') - \Lambda_{\mathrm{X}}'(x')\right|$$
$$+ \sup_{|x-x'|\le A_2h_0}\left|\tilde{\Lambda}_1''(x') - \Lambda_{\mathrm{X}}''(x')\right| + \sup_{|x-x'|\le A_2h_0}\left|\tilde{\Lambda}_1'''(x') - \Lambda_{\mathrm{X}}'''(x')\right|.$$

We will now prove that all three terms on the RHS above are $O_p\left(h_2^2\right)$. For example,

$$\sup_{|x-x'|\le A_2h_0}\left|\tilde{\Lambda}_1'(x') - \Lambda_{\mathrm{X}}'(x')\right| = \sup_{|x-x'|\le A_2h_0}\left|\hat{\lambda}_{T,1}(x') - \lambda_X(x')\right| \le$$
$$\sup_{|x-x'|\le A_2h_0}\left|\hat{\lambda}_{T,1}(x') - \mathbb{E}\hat{\lambda}_{T,1}(x')\right| + \sup_{|x-x'|\le A_2h_0}\left|\mathbb{E}\hat{\lambda}_{T,1}(x') - \lambda_X(x')\right| = O_P\left(h_1^2\right)$$

by lemma 4.3.2 and the standard result that the asymptotic bias of $\hat{\lambda}_{T,1}(x')$ is proportional to $h_1^2\lambda_Y''(x')$. The other two terms are handled in exactly the same way and give the same result. Hence, for $l = 1,2,3$ we have

$$\sup_{|x-x'|\le A_2h_0}\left|\tilde{\Lambda}_1^{(l)}(x') - \Lambda_{\mathrm{X}}^{(l)}(x')\right| = O_p\left(h_1^2\right),$$

and since $h_1/h_2 \to c_3$ as $n \to +\infty$, $c_3 > 0$ a constant, we conclude that

$$\sup_{|x-x'|\le A_2h_0}\sum_{m=1}^{3}\left|g^{(m)}(x') - \Lambda_{\mathrm{X}}^{(m)}(x')\right| = O_p\left(h_2^2\right). \quad \blacksquare$$

Lemma 4.3.4. *Let $\hat{\Lambda}_1 \equiv \hat{\Lambda}_{1,\mathrm{X}}(x)$ and $\tilde{\Lambda}_1$ be as defined on pages 71 and 72 respectively. Then, under the conditions of theorem 4.3.1*

(i)

$$\sup_{|x-x'|\leq A_2 h_0} \left| \hat{\Lambda}_1'(x') - \tilde{\Lambda}_1'(x') \right| = O_p(h_2^4)$$

(ii)

$$\sup_{|x-x'|\leq A_1 h_0} \left| \hat{\Lambda}_1(x') - \tilde{\Lambda}_1(x') \right| = O_p(h_2^5).$$

Proof. We start with (i). Expanding $\hat{\Lambda}_1'$ in Taylor series around x' with the Cauchy form of the remainder gives

$$\hat{\Lambda}_1'(x') = \hat{\Lambda}_1'(x) + \hat{\Lambda}_1''(x)(x'-x) + \hat{\Lambda}_1'''(x)\frac{(x'-x)^2}{2} + \hat{\Lambda}_1^{(4)}(x)\frac{(x'-x)^3}{3!} + \hat{\Lambda}_1^{(5)}(\xi)\frac{(x'-x)^4}{4!}$$

for some ξ in (x, x'). Differentiating $\tilde{\Lambda}_1$ with respect to x' gives

$$\tilde{\Lambda}_1'(x') = \hat{\Lambda}_1'(x) + \hat{\Lambda}_1''(x)(x'-x) + \hat{\Lambda}_1'''(x)\frac{(x'-x)^2}{2} + \hat{\Lambda}_1^{(4)}(x)\frac{(x'-x)^3}{3}.$$

We immediately see that

$$\hat{\Lambda}_1'(x') - \tilde{\Lambda}_1'(x') = \hat{\Lambda}_1^{(5)}(\xi)\frac{(x'-x)^4}{4!}.$$

Now,

$$\left| \hat{\Lambda}_1^{(5)}(\xi) \right| \leq \sup_{|x-x'|\leq A_2 h_0} \left| \hat{\Lambda}_1^{(5)}(x') \right|$$

and from the conditions of the theorem we have that $h_2/h_0 \to c_2$, thus

$$\sup_{|x-x'|\leq A_2 h_0} \left| \hat{\Lambda}_1'(x') - \tilde{\Lambda}_1'(x') \right| \leq \sup_{|x-x'|\leq A_2 h_0} \left| \hat{\Lambda}_1^{(5)}(x') \right| (A_2 h_0)^4 = O_p(h_2^4).$$

To prove (ii), notice that

$$\int \left(\hat{\Lambda}_1'(x') - \tilde{\Lambda}_1'(x') \right) \, dx' = \hat{\Lambda}_1(x') - \tilde{\Lambda}_1(x') + c.$$

Therefore,

$$\begin{aligned} \sup_{|x-x'|\leq A_1 h_0} \left| \hat{\Lambda}_1(x') - \tilde{\Lambda}_1(x') \right| &\leq \sup_{|x-x'|\leq A_1 h_0} \left| \hat{\Lambda}_1^{(5)}(\xi) \right| \int \frac{(x'-x)^4}{4!} \, dx' \\ &\leq \sup_{|x-x'|\leq A_1 h_0} \left| \hat{\Lambda}_1^{(5)}(x') \frac{(x'-x)^5}{5!} \right| = O_p(h_2^5). \quad \blacksquare \end{aligned}$$

4.4 Implementation of $\hat{\lambda}_{n,2}(x|h_1, h_2)$ and $\hat{\lambda}_{T,2}(x|h_1, h_2)$.

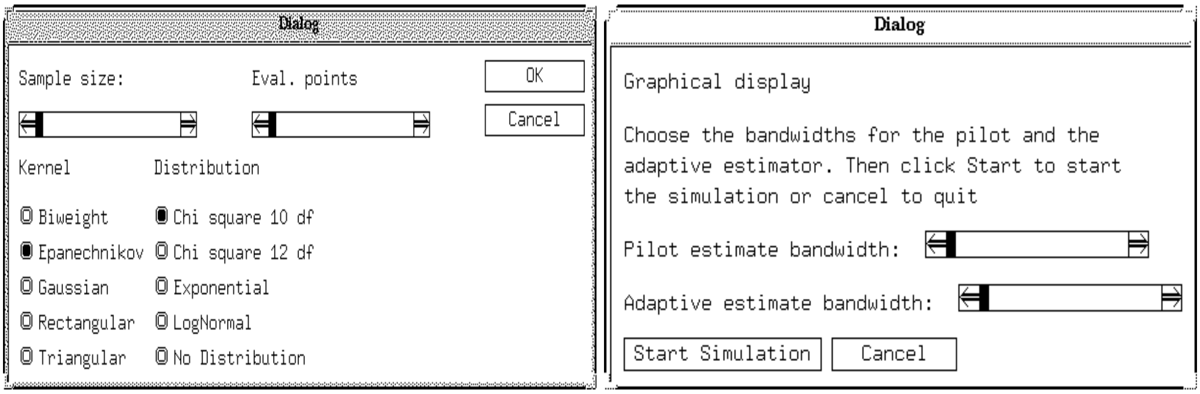
The main task in this section is to implement estimators $\hat{\lambda}_{n,2}$ defined in chapter 3 and the transformed estimate $\hat{\lambda}_{T,2}$, so that we can examine their visual appearance. In subsection 4.4.1 we develop a utility that plots the estimators by allowing interactive choice of bandwidths and method of estimation. The resulting estimates are drawn overlayed on the same window. Motivation comes from the fact that on an interactive 'trial and error' process, the user gets substantial insight on the behavior of the estimator for different bandwidth values ([39], page 44). Comparison of the estimators is very difficult as the estimators are not at the same footing (level). For this reason, in subsection 4.4.2 we use this utility to illustrate the above estimators on distributional data.

4.4.1 Interactive bandwidth choice.

In this subsection we present an interactive method for choosing the bandwidth of estimators $\hat{\lambda}_{n,2}$ and $\hat{\lambda}_{T,2}$. We implement a Graphical User Interface (GUI) so that the user can see the behavior of the estimator for the current bandwidth and compare it with previous choices. To implement the method on a computer, we have used XLisp-Stat (Tierney, [46]), an object-oriented programming language which provides dynamical graphical mechanisms and convenient tools for drawing the estimates.

From the user's point of view, interaction with the system (and thus control of the estimators) is driven by 'dialogs' which are special forms of programming objects. Their purpose is to provide an easy way for the user to specify any parameters to be used. In our case this would be the kernel, the bandwidth etc. Within a dialog there are two ways for the user to choose the parameters of his choice depending on the kind of the parameter to be selected. If the user wants to choose, say, the sample size then this is done by clicking the arrows of a 'slider'. On the other hand if the user wants to specify, say, the distribution to be used, this is done simply by ticking the corresponding 'circle' from the list of available distributions.

We have organized the process of selection of parameters in two stages with each stage being a dialog. As can be seen in figure 4.1(a) at first stage the user specifies general parameters, i.e. distribution, sample size, etc.If the option 'No Distribution' is selected, a new window pops up prompting the user to specify the file that contains the raw, univariate data separated by spaces in list form.The second stage (figure 4.1(b)) is selection of estimation method (variable bandwidth or transformation) and bandwidth values. This second dialog is present throughout the estimation procedure so that the user can each time choose a different estimation method or bandwidths. The output appears on a third window after the user clicks 'OK' for the first time. It has the true hazard rate (if distributional data is used) and the estimate. Choosing new bandwidth values and/or another estimation method and clicking OK draws the new estimate, overlayed on the output window. Such a window can be seen in figures 4.2-4.4. As can be seen from the pictures, every estimate is drawn with a different color. In the current implementation every new estimate up to the seventh appears on the screen with different color and beyond the seventh, every new estimate is drawn in black. Finally the procedure stops at any time simply by clicking 'Cancel' on the second dialog.



(a) dialog 1 (b) dialog 2

Figure 4.1: Dialog windows for interactive bandwidth choice.

4.4.2 Implementation of the estimators.

With the utility described in the previous subsection, we illustrate the transformed estimate, $\hat{\lambda}_{T,2}$, and the adaptive estimator, $\hat{\lambda}_{n,2}$. The distributions chosen are χ^2_{12}, $LN(0,1)$ and the mixture of normals $\frac{2}{3}N(2,1)+\frac{1}{3}N(3,.2^2)$. In all cases we use the exponential target, transforming samples of magnitude $n=1000$ to exponential samples with the transformation

$$g(x)=-\log(1-\hat{F}_X(x)),\ \ \hat{F}_X(x)=\sum_{i=1}^{n}\mathcal{K}\left(\frac{x-X_i}{h}\right)$$

where $\mathcal{K}$ is the integrated epanechnikov kernel. In all figures the bandwidths for the adaptive have been chosen using the Silverman's default bandwidth method [39]. For the transformed estimate we use the method of Ruppert and Cline [35], i.e. the bandwidth is calculated by multiplying the interquartile range of the data with a bandwidth factor.

In general from figures 4.2 and 4.4, it seems that the adaptive and the transformed estimates have quite similar behavior for some hazard rates. That was the case for other distributions, such as the uniform, which are not illustrated here. An explanation for this can be given by the connection in the structure of the two estimates as this was discussed in subsection 1.5.2. In figure 4.2 we estimate the hazard rate of the χ^2_{12} distribution. Here as bandwidths we used $h_1=1.26$ for the pilot of the adaptive and $h_2=1.07$ for $\hat{\lambda}_{n,2}$. The bandwidth factor for the transformed estimator is .2. As we see, the adaptive behaves very poorly at the left boundary but its performance from $x=7$ and on is fairly similar to that of the transformed with the second one being slightly better. The poor behavior of all estimators from $x=20$ and on is due to lack of data and to an increase in the variance of the estimator for larger x. For this reason we will not asses their performance beyond that point.

In figure 4.3 we asses the performance of the estimators in the case of standard lognormal hazard rate. We used bandwidth $h_1=.253$ for the pilot of the adaptive, $h_2=.215$ for the adaptive itself and for $\hat{\lambda}_{T,2}$ the bandwidth factor is .05. The performance of both estimators in this case is not as good as previously, however, from $x=0$ to $x=1.2$

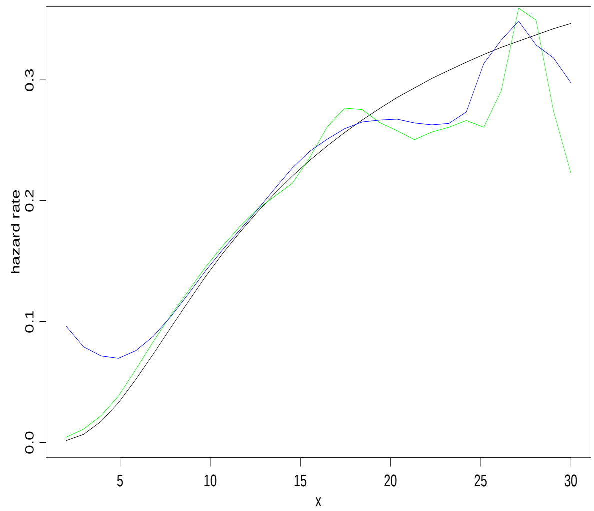


Figure 4.2: Estimation of the hazard rate of χ^2_{12} distribution (black line) by $\hat{\lambda}_{T,2}$ (green) and $\hat{\lambda}_{n,2}$ (blue).

the transformed estimator has a decent behavior especially compared to the behavior of the adaptive. From $x = 1.2$ and on the adaptive performs visibly better.

In figure 4.4 we use a mixture of two normals to examine the performance of the estimators. The bandwidths we used are $.2$ for both bandwidths of the adaptive and bandwidth factor $.12$ for the transformed. The conclusion from this picture is that both estimators have similar behavior, with the transformed estimate to be doing a little better than the adaptive and that apart from the problems that both estimators seem to have from about $x = 3.4$ and on, their performance can be described as satisfactory.

In figure 4.5 we use the transformed and the adaptive estimates with the suicide data which were also analyzed in [39]. First, all estimates suggest that there are two peaks at about 100 and 230 and a third peak at about $x = 600$. Both the transformed and the adaptive suggest two smaller modes near 100. The qualitative interpretation though is the same for all estimates and is in accordance with the results of [39]. The hazard from $x = 100$ to $x = 400$ is diminishing but not a constant rate; An initial fall in hazard is followed by a short uplift at the second peak. The curve is then dropping again but a slower rate. The increase at the third peak can be attributed to the sparse data at that region and thus can be regarded as random effect. The green line is the true failure rate, the blue is the adaptive, the red is the pilot and the black is the ideal.

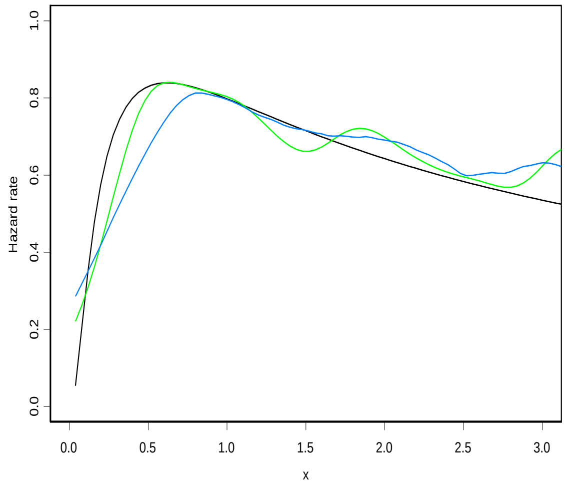


Figure 4.3: Estimation of the hazard rate of the $LN(0,1)$ distribution (black line) by $\hat{\lambda}_{T,2}$ (green) and $\hat{\lambda}_{n,2}$ (blue).

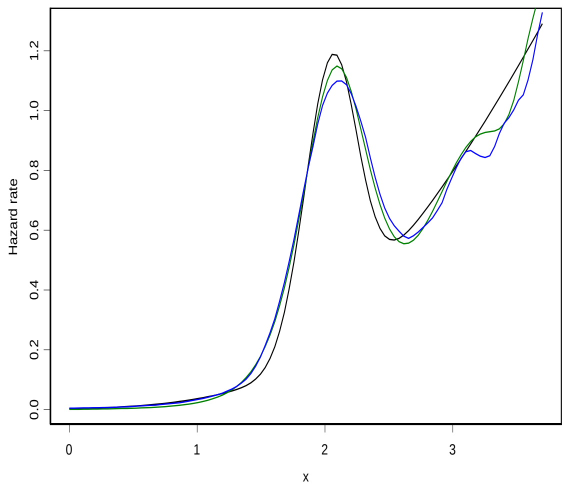


Figure 4.4: Estimation of the hazard rate of the $\frac{2}{3}N(2,1) + \frac{1}{3}N(3,.2^2)$ distribution (black line) by $\hat{\lambda}_{T,2}$ (green) and $\hat{\lambda}_{n,2}$ (blue).

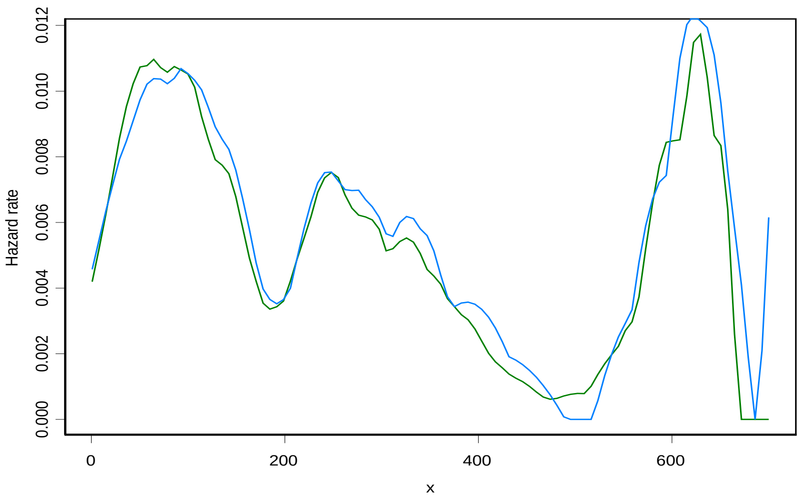


Figure 4.5: Estimation of the hazard rate by the transformed (green line) and the adaptive (blue line) estimates for the suicide data.

Chapter 5

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

5.1 Concluding remarks.

This work began with the aim to develop kernel based hazard rate estimators that perform better than traditional hazard rate estimates such as those defined in the introduction. Identifying the problems of bias and boundary bias as two areas that there is room for improvement, we have provided the methodologies to alleviate these problems.

In particular, in chapter 2 we showed that a local linear based estimator of the hazard rate treats effectively the frequently met boundary bias problem of conventional kernel hazard rate estimators. As bandwidth choice is of particular significance for the accuracy of kernel estimators, the two bandwidth choice procedures proposed in chapter 2 lead to optimal performance of the local linear estimator.

In chapter 3, extending the ideas of Hall and Marron [16] for estimation location variable bandwidth to the hazard rate case, we applied the square root law to two of the most popular fixed bandwidth kernel based estimates of the hazard rate with the purpose of improving precision. We showed that the adaptive estimators, achieve a faster bias rate of convergence compared to the bias rate of fixed bandwidth estimators. Comparing the adaptive estimates with their 'ideal' counterparts so that we have a measure of the heights that the adaptive estimates might aspire, we showed that the performance of the adaptive estimates is almost, but not quite, as good as the performance of the ideals. Furthermore, the distance between the ideal and the adaptive estimate in each case is found to be a random variable, asymptotically normally distributed.

In chapter 4 we extend the method of empirical transformations which was applied to density settings by Ruppert and Cline [35]. The asymptotical square error of the resulting estimator shows that the improvement in this case is of the same magnitude as in the variable bandwidth approach. We then examine the performance of the transformed estimate with the one of the adaptive estimator via simulations on distributional data. The resulting plots indicate that although their behavior is similar, the transformed estimate behaves marginally better.

In a nutshell, using the estimators developed herein in an optimal fashion can lead to significant improvements over traditional kernel hazard rate estimators. However, there exist cases where the developed estimates, even when employed optimally, will not be supported by the data and thus will yield inferior estimates. Given that it is doubtful that a globally best way of estimation exists, we feel that what is needed is a variety of estimation procedures with clear understanding of their strengths and weaknesses.

5.2 Ideas for future work.

In many real world applications of the hazard rate, the available samples are censored. Therefore a natural step forward would be to extent the methods studied in this thesis to the censored case. We proceed now with individual suggestions for each chapter.

As seen in chapter 2, the rate of convergence of the bias of the local linear fit estimate developed there is of the same order as of that of traditional kernel hazard rate estimates. Therefore, we expect that an estimator will inherit the advantages of both approaches if we combine the ideas of variable bandwidth and local linear fit. That is, by using the bandwidth law used for estimator $\hat{\lambda}_{n,2}$ in estimator $\hat{\lambda}_L$ we expect to get an estimator with the same bias rate of convergence as that of $\hat{\lambda}_{n,2}$ but without the boundary effects it produces. As pointed out in subsections 2.5.1 and 2.5.2 one of the top priorities will be to study asymptotic properties of the proposed bandwidth choice methods.

Much of the work in chapter 3 was devoted to the study of the asymptotic properties of the estimators developed there. As we have seen, bandwidth choice plays an important role in the performance of kernel based estimators, thus bandwidth selection for the adaptive estimator $\hat{\lambda}_{n,2}$ is a significant future research topic. In section 4.4 we saw that the plots suggest a similarity in the performance of estimators $\hat{\lambda}_{n,2}$ and $\hat{\lambda}_{T,2}$. It will be interesting to see if this similarity is true in a proper comparison of the estimators. Furthermore such a comparison will yield valuable insight to the practical performance of the estimators and will provide a measure of preference as to which one is better under certain circumstances.

In chapter 4 we confined ourselves to the study of the asymptotic squared error of estimator $\hat{\lambda}_{T,2}$. A natural suggestion thus is to precise formulation of the asymptotic bias of this estimator. Furthermore, it is known that asymptotic normality of an estimator is important because it allows the construction of confidence intervals which are very useful in practical fields. Thus establishing the asymptotic normality of $\hat{\lambda}_{T,2}$ is of particular interest for applications of the estimator.

Appendix A

CALCULATIONS

A.1 Calculations, chapter 2.

A.1.1 Minimization - local linear fit.

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \{Y_i - \beta_0 - \beta_1(x - x_i)\}^2 K_h(x - x_i) = \min_{\beta_0,\beta_1} SS$$

If we call $\hat{\beta}_0, \hat{\beta}_1$ the solution of the least squares problem then the estimator of the hazard rate is clearly $\hat{\beta}_0(x)$. Minimization of the above sum of squares is achieved by solving the following system of equations

$$\frac{\partial SS}{\partial \beta_0} = 0, \quad \frac{\partial SS}{\partial \beta_1} = 0.$$

We have

$$\frac{\partial SS}{\partial \beta_0} = -2\sum_{i=1}^{n} \left\{Y_i - \hat{\beta}_0 - \hat{\beta}_1(x - x_i)\right\} K_h(x_i - x) = 0$$
$$\frac{\partial SS}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} \left\{Y_i - \hat{\beta}_0 - \hat{\beta}_1(x - x_i)\right\} K_h(x_i - x)(x - x_i) = 0$$

Rearranging we get

$$\sum_{i=1}^{n} Y_i K_h(x_i - x) = \hat{\beta}_0 \sum_{i=1}^{n} K_h(x_i - x) + \hat{\beta}_1 \sum_{i=1}^{n} K_h(x_i - x)(x - x_i) \tag{A.1}$$

$$\sum_{i=1}^{n} Y_i K_h(x_i - x)(x - x_i) = \hat{\beta}_0 \sum_{i=1}^{n} K_h(x_i - x)(x - x_i) + \hat{\beta}_1 \sum_{i=1}^{n} K_h(x_i - x)(x - x_i)^2 \tag{A.2}$$

Solving (A.1) for $\hat{\beta}_1$ yields

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i K_h(x_i - x) - \hat{\beta}_0 \sum_{i=1}^{n} K_h(x_i - x)}{\sum_{i=1}^{n} K_h(x_i - x)(x - x_i)}$$

substitute $\hat{\beta}_1$ to (A.2) to get

$$\sum_{i=1}^{n} Y_i K_h(x_i - x)(x - x_i) = \hat{\beta}_0 \sum_{i=1}^{n} K_h(x_i - x)(x - x_i) +$$
$$\sum_{i=1}^{n} K_h(x_i - x)(x - x_i)^2 \frac{\sum_{i=1}^{n} Y_i K_h(x_i - x) - \hat{\beta}_0 \sum_{i=1}^{n} K_h(x_i - x)}{\sum_{i=1}^{n} K_h(x_i - x)(x - x_i)}$$

rearranging gives

$$\sum_{i=1}^{n} Y_i K_h(x_i - x)(x - x_i) - \frac{\sum_{i=1}^{n} Y_i K_h(x_i - x) \sum_{i=1}^{n} K_h(x_i - x)(x - x_i)^2}{\sum_{i=1}^{n} K_h(x_i - x)(x - x_i)} =$$
$$= \frac{\hat{\beta}_0 \left(\sum_{i=1}^{n} K_h(x_i - x)(x - x_i) \right)^2 - \hat{\beta}_0 \sum_{i=1}^{n} K_h(x_i - x) \sum_{i=1}^{n} K_h(x_i - x)(x - x_i)^2}{\sum_{i=1}^{n} K_h(x_i - x)(x - x_i)}$$

We can rewrite the above equation as

$$\frac{\left(\sum_{i=1}^{n} Y_i K_h(x_i - x)(x - x_i) \right)^2 - \sum_{i=1}^{N} Y_i K_h(x_i - x)(x - x_i) \sum_{i=1}^{n} K_h(x_i - x)(x - x_i)^2}{\sum_{i=1}^{n} K_h(x_i - x)(x - x_i)}$$
$$= \frac{\hat{\beta}_0 \left(\left(\sum_{i=1}^{n} K_h(x_i - x)(x - x_i) \right)^2 - \sum_{i=1}^{n} K_h(x_i - x) \sum_{i=1}^{n} K_h(x_i - x)(x - x_i)^2 \right)}{\sum_{i=1}^{n} K_h(x_i - x)(x - x_i)}$$

Then solving for $\hat{\beta}_0$ gives

$$\hat{\beta}_0 =$$
$$\frac{\sum_{i=1}^{n} Y_i K_h(x_i - x)(x - x_i) \sum_{i=1}^{n} K_h(x_i - x)(x - x_i) - \sum_{i=1}^{n} Y_i K_h(x_i - x) \sum_{i=1}^{n} K_h(x_i - x)(x - x_i)^2}{\left(\sum_{i=1}^{n} Y_i K_h(x_i - x)(x - x_i) \right)^2 - \sum_{i=1}^{n} Y_i K_h(x_i - x)(x - x_i) \sum_{i=1}^{n} K_h(x_i - x)(x - x_i)^2}$$

A.1.2 Proof of lemma 2.4.1.

The proof was outlined originally in [7], pp. 30-31. Here we give a somewhat more detailed outline of the original proof.

Proof. Writing the difference between

$$\sum_{i=1}^{n} G(t_i)B \text{ and } \int G(s)\,ds$$

as a sum of integrals gives

$$\left|\sum_{i=1}^{n} G(t_i)B - \int G(s)\,ds\right| = \left|\sum_{i=1}^{n} G(t_i)B - \sum_{i=1}^{n}\int_{t_i-\frac{B}{2}}^{t_i+\frac{B}{2}} G(s)\,ds\right| =$$
$$\left|\sum_{i=1}^{n}\int_{t_i-\frac{B}{2}}^{t_i+\frac{B}{2}} \{G(t_i)-G(s)\}\,ds\right|.$$

Now, taking the second order Taylor expansion of $G(t_i)$ and $G(s)$ around x_i with the Lagrange form of the remainder which is

$$G(t_i) = G(x_i) + (t_i - x_i)G'(x_i) + \int_{x_i}^{t_i} G''(t)(t_i - t)\,dt$$
$$G(s) = G(x_i) + (s - x_i)G'(x_i) + \int_{x_i}^{s} G''(t)(s - t)\,dt$$

and subtracting, gives

$$G(t_i) - G(s) = (t_i - s)G'(x_i) + \int_{x_i}^{t_i} G''(t)(t_i - t)\,dt - \int_{x_i}^{s} G''(t)(s - t)\,dt$$

$$= (t_i - s)G'(x_i) + t_i(G'(t_i) - G'(x_i)) - s(G'(s) - G'(x_i)) + \int_{t_i}^{s} G''(t)t\,dt$$
$$= G'(t_i)(t_i - s) + sG'(t_i) - sG'(s) + \int_{t_i}^{s} tG''(t)\,dt.$$

Thus,

$$G(t_i) - G(s) = G'(t_i)(t_i - s) + \int_{s}^{t_i} (s - t)G''(t)\,dt. \tag{A.3}$$

Now,

$$\int_{s}^{t_i} (s - t)G''(t)\,dt = G'(t_i)(s - t_i) + \int_{s}^{t_i} G''(t)\,dt. \tag{A.4}$$

Putting (A.4) back in (A.3) gives

$$G(t_i) - G(s) = \int_{s}^{t_i} G''(t).$$

Notice that

$$\int_{t_i-\frac{B}{2}}^{t_i+\frac{B}{2}} G'(t_i)(t_i - s)\,ds = 0.$$

Thus, the integral of the difference $G(t_i) - G(s)$ can be written as

$$\int_{t_i-\frac{B}{2}}^{t_i+\frac{B}{2}} (G(t_i) - G(s))\, ds = \int_{t_i-\frac{B}{2}}^{t_i+\frac{B}{2}} \left\{ G'(t_i)(t_i - s) + \int_s^{t_i} G''(t)\, dt \right\} ds.$$

Note that

$$G'(t_i)(t_i - s) = \int_s^{t_i} (t-s)G''(t)\, dt + \int_{t_i}^{s} G''(t)\, dt = \int_s^{t_i} (t-s)G''(t)\, dt - \int_s^{t_i} G''(t)\, dt.$$

Now, let

$$D = \left\{ (s,t) : t_i - \frac{B}{2} \le s \le t_i + \frac{B}{2} \text{ and } s \le t \le t_i \right\} = D_1 + D_2$$

where

$$D_1 = \left\{ (s,t) : t_i - \frac{B}{2} \le s \le t_i,\ s \le t \le t_i \right\}, D_2 = \left\{ (s,t) : t_i \le s \le t_i + \frac{B}{2},\ s \le t \le t_i \right\}.$$

Then,

$$\int_{t_i-\frac{B}{2}}^{t_i+\frac{B}{2}} \left\{ G'(t_i)(t_i - s) + \int_s^{t_i} G''(t)\, dt \right\} ds = \iint_D (t-s)G''(t)\, dt\, ds =$$

$$\left\{ \iint_{D_1} + \iint_{D_2} \right\} (t-s)G''(t)\, dt\, ds = \iint_{D_1} (t-s)G''(t)\, dt\, ds - \iint_{D_3} (t-s)G''(t)\, dt\, ds.$$

where

$$D_3 = \left\{ (s,t) : t_i \le s \le t_i + \frac{B}{2},\ t_i \le t \le s \right\}.$$

Changing the order of integration in D_1, the new bounds for s and t will be

$$t_i - \frac{B}{2} \le s \le t_i,\ s \le t \le t_i \Rightarrow t_i - \frac{B}{2} \le t \le t_i,\ \ t_i - \frac{B}{2} \le s \le t.$$

Working similarly in D_3, the new areas of integration become

$$D_4 = \left\{ (s,t) : t_i - \frac{B}{2} \le t \le t_i,\ t_i - \frac{B}{2} \le s \le t \right\}, D_5 = \left\{ (s,t) : t_i \le t \le t_i + \frac{B}{2},\ t \le s \le t_i + \frac{B}{2} \right\}.$$

Then,

$$\iint_D (t-s)G''(t)\, dt\, ds = \iint_{D_4} (t-s)G''(t)\, ds\, dt - \iint_{D_5} (t-s)G''(t)\, ds\, dt.$$

Denote with $||s-t||$ as the maximal difference between s and t. We have,

$$\int_{t_i-\frac{B}{2}}^{t_i} \int_{t_i-\frac{B}{2}}^{t} (s-t)G''(t)\, ds\, dt \le ||s-t|| \int_{t_i-\frac{B}{2}}^{t_i} \int_{t_i-\frac{B}{2}}^{t} G''(t)\, ds\, dt \le ||s-t|| \frac{B}{2} \int_{t_i-\frac{B}{2}}^{t_i} G''(t)\, dt.$$

Also,

$$\int_{t_i}^{t_i+\frac{B}{2}}\int_{t}^{t_i+\frac{B}{2}}(s-t)G''(t)\,dt\,ds \leq ||s-t||\int_{t_i}^{t_i+\frac{B}{2}}\int_{t}^{t_i+\frac{B}{2}}G''(t)\,dt\,ds \leq ||s-t||\frac{B}{2}\int_{t_i}^{t_i+\frac{B}{2}}G''(t)\,dt.$$

Therefore,

$$\iint_{D_4}(t-s)G''(t)\,ds\,dt - \iint_{D_5}(t-s)G''(t)\,ds\,dt =$$

$$||s-t||\frac{B}{2}\left(\int_{t_i-\frac{B}{2}}^{t_i}G''(t)\,dt - \int_{t_i}^{t_i+\frac{B}{2}}G''(t)\,dt\right) \leq \frac{B^2}{4}\left(\int_{t_i-\frac{B}{2}}^{t_i}G''(t)\,dt + \int_{t_i}^{t_i+\frac{B}{2}}G''(t)\,dt\right)$$
$$= \frac{B^2}{4}\int_{t_i-\frac{B}{2}}^{t_i+\frac{B}{2}}G''(t)\,dt$$

and therefore,

$$\left|\int_{t_i-\frac{B}{2}}^{t_i}\int_{t_i-\frac{B}{2}}^{t}(s-t)G''(t)\,ds\,dt - \int_{t_i}^{t_i+\frac{B}{2}}\int_{t}^{t_i+\frac{B}{2}}(s-t)G''(t)\,dt\,ds\right| \leq \frac{B^2}{4}\int_{t_i-\frac{B}{2}}^{t_i+\frac{B}{2}}|G''(t)|\;dt.$$

Since

$$\sum_{i=1}^{n}\int_{t_i-\frac{B}{2}}^{t_i+\frac{B}{2}}|G''(t)|\;dt = \int |G''(t)|\;dt$$

it follows that

$$\left|\sum_{i=1}^{n}G(t_i)B - \int G(s)\,ds\right| \leq \frac{B^2}{4}\int |G''(t)|\;dt. \qquad \blacksquare$$

A.1.3 Functionals of the hazard rate.

Assuming that all the derivatives involved exist, and with

$$w(x) = \begin{cases} 1 & x \in [0,T] \\ 0 & \text{elsewhere} \end{cases}$$

we have

$$\begin{aligned} R\left(\lambda^{(s)}(x)\right) &= \int \lambda^{(s)}(x)\lambda^{(s)}(x)w(x)\,dx \\ &= \lambda^{(s)}(x)\lambda^{(s-1)}(x)w(x) - \int \lambda^{(s+1)}(x)\lambda^{(s-1)}(x)w(x)\,dx \end{aligned}$$

Therefore,

$$\begin{aligned}
R\left(\lambda^{(s)}(x)\right) &= -\int \lambda^{(s+1)}(x)\lambda^{(s-1)}(x)w(x)\,dx \\
&= \lambda^{(s+1)}(x)\lambda^{(s-2)}(x)w(x) + \int \lambda^{(s+2)}(x)\lambda^{(s-2)}(x)w(x)\,dx \\
&= \int \lambda^{(s+2)}(x)\lambda^{(s-2)}(x)w(x)\,dx \\
&\ldots\ldots\ldots \\
&= (-1)^s \int \lambda^{(2s)}(x)\lambda(x)w(x)\,dx.
\end{aligned}$$

Note that we used

$$\lambda^{(s)}(x)\lambda^{(s-1)}(x)w(x) = \lambda^{(s+1)}(x)\lambda^{(s-2)}(x)w(x) = \cdots = \lambda^{(2s-1)}(x)\lambda'(x)w(x) = 0$$

since for x outside $(0, T)$, we have $w(x) = 0$.

A.1.4 Estimation of ψ_r with reference to $W(\beta, \kappa)$.

For a Weibul distribution with scale parameter β and index κ the rth derivative of the hazard rate is

$$\lambda^{(r)}(x) = \kappa^{r+1}\beta(\beta-1)(\beta-2)(\beta-r)(\kappa x)^{\beta-r-1}.$$

Then,

$$\begin{aligned}
\psi_r &= \int_0^b \lambda^{(r)}(t)\lambda(t)\,dt \\
&= \int_0^b \kappa^{r+1}\beta(\beta-1)(\beta-2)\ldots(\beta-r)(\kappa t)^{\beta-r-1}\kappa\beta(\kappa t)^{\beta-1}\,dt \\
&= \kappa^{r+2}\beta^2(\beta-1)(\beta-2)\ldots(\beta-r)\int_0^b (\kappa t)^{2\beta-r-2}\,dt \\
&= \kappa^{2\beta}\beta^2(\beta-1)(\beta-2)\ldots(\beta-r)\int_0^b t^{2\beta-r-2}\,dt \\
&= \frac{\kappa^{2\beta}\beta^2(\beta-1)\ldots(\beta-r)}{2\beta-r-1}b^{2\beta-r-1}.
\end{aligned}$$

A.1.5 Mean and variance of $\hat{\psi}_r(g)$.

Let

$$w_{i,j} = w(X_{(i)})w(X_{(j)}), \quad w_{i,j,k} = w(X_{(i)})w(X_{(j)})w(X_{(k)}).$$

Lemma A.1.1. *Let $X_1, \ldots, X_n$ be a sample from some density f having cdf F. Then, under the assumptions of theorem 2.5.1*

$$\mathbb{E}\hat{\psi}'_r(g) = \iint L_g^{(r)}(x-y)\lambda(x)\lambda(y)\,dx\,dy.$$

Proof. The proof of the result for estimator $\hat{\psi}_r'(g)$ is trivial, and so we will prove the same result for estimator $\hat{\psi}_r(g)$.

$$\mathbb{E}\hat{\psi}_r(g) = \mathbb{E}\left(\sum\sum_{i\neq j} \frac{L_g^{(r)}\left(X_{(i)} - X_{(j)}\right)}{(n-j+1)(n-i+1)} w_{i,j}\right) =$$

$$\mathbb{E}\left\{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(j)}\right) w_{i,j}}{(n-j+1)(n-i+1)} + \sum_{j=1}^{n-1}\sum_{i=j+1}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(j)}\right) w_{i,j}}{(n-j+1)(n-i+1)}\right\} \tag{A.5}$$

Write the first double sum of the mean as

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(j)}\right) w_{i,j}}{(n-j+1)(n-i+1)} =$$

$$\sum_{j=2}^{n} \frac{L_g^{(r)}\left(X_{(1)} - X_{(j)}\right) w_{1,j}}{n(n-j+1)} + \sum_{j=3}^{n} \frac{L_g^{(r)}\left(X_{(2)} - X_{(j)}\right) w_{2,j}}{(n-j+1)(n-1)} + \sum_{j=4}^{n} \frac{L_g^{(r)}\left(X_{(3)} - X_{(j)}\right) w_{3,j}}{(n-j+1)(n-2)} +$$

$$\cdots + \sum_{j=n-1}^{n} \frac{L_g^{(r)w_{1,j}}\left(X_{(n-2)} - X_{(j)}\right) w_{n-2,j}}{3(n-j+1)} + \frac{L_g^{(r)}\left(X_{(n-1)} - X_{(n)}\right) w_{n-1,j}}{2\cdot 1}$$

and the second one as

$$\sum_{j=1}^{n-1}\sum_{i=j+1}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(j)}\right) w_{i,j}}{(n-j+1)(n-i+1)} = \sum_{i=2}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(1)}\right) w_{i,1}}{n(n-i+1)} + \sum_{i=3}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(2)}\right) w_{i,2}}{(n-1)(n-i+1)}$$

$$\sum_{i=4}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(3)}\right) w_{i,3}}{(n-2)(n-i+1)} + \cdots + \sum_{i=n-1}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(n-2)}\right) w_{i,n-2}}{3(n-i+1)}$$

$$+ \frac{L_g^{(r)}\left(X_{(n)} - X_{(n-1)}\right) w_{n,n-1}}{2\cdot 1}$$

We will examine each one of these terms individually. As known the joint distribution of two order statistics $X_{(i)}$ and $X_{(j)}$, $i < j$ is

$$f(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F(x)^{i-1}(F(y) - F(x))^{j-i-1}(1 - F(y))^{n-j} f(x) f(y)$$

We will apply this formula to both terms above. Starting with he second term of (A.5), set

$$\sum_{j=2}^{n} \frac{L_g^{(r)}\left(X_{(1)} - X_{(j)}\right) w_{1,j}}{n(n-j+1)} = \hat{\psi}_{r,1}, \quad \sum_{j=3}^{n} \frac{L_g^{(r)}\left(X_{(2)} - X_{(j)}\right) w_{2,j}}{(n-j+1)(n-1)} = \hat{\psi}_{r,2}, \ldots,$$

$$\frac{L_g^{(r)}\left(X_{(n-1)} - X_{(n)}\right) w_{n-1,n}}{1\cdot 2} = \hat{\psi}_{r,n-1}$$

and

$$\sum_{i=2}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(1)}\right) w_{i,1}}{(n-1+1)(n-i+1)} = \hat{\psi}'_{r,2}, \quad \sum_{i=3}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(2)}\right) w_{i,2}}{(n-2+1)(n-i+1)} = \hat{\psi}'_{r,3}, \quad \ldots$$

$$\ldots \quad \frac{L_g^{(r)}\left(X_{(n)} - X_{(n-1)}\right) w_{n,n-1}}{2(n-i+1)} = \hat{\psi}'_{r,n}.$$

Obviously

$$\sum_{i=1}^{n-1} \mathbb{E}\hat{\psi}_{r,i} = \mathbb{E}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(j)}\right) w_{i,j}}{(n-j+1)(n-i+1)}$$

and

$$\sum_{i=2}^{n} \mathbb{E}\hat{\psi}'_{r,i} = \mathbb{E}\sum_{j=1}^{n-1}\sum_{i=j+1}^{n} \frac{L_g^{(r)}\left(X_{(i)} - X_{(j)}\right) w_{i,j}}{(n-j+1)(n-i+1)}$$

and therefore

$$\mathbb{E}\hat{\psi}_r = \sum_{i=1}^{n-1} \mathbb{E}\hat{\psi}_{r,i} + \sum_{i=2}^{n} \mathbb{E}\hat{\psi}'_{r,i}.$$

Then, integrating over the set $D_1 = \{(x,y) : x < y\}$, we have

$$\begin{aligned}
\mathbb{E}\hat{\psi}_{r,1} = \frac{1}{n}\iint \sum_{j=2}^{n} L_g^{(r)}(x-y)\frac{n(n-1)!}{(j-2)!(n-j+1)!}(F(y)-F(x))^{j-2}(1-F(y))^{n-j} \\
\times f(x)f(y)\,dx\,dy
\end{aligned}$$

$$\begin{aligned}
&= \iint \sum_{j=2}^{n} L_g^{(r)}(x-y)\binom{n-1}{j-2}\left(\frac{F(y)-F(x)}{1-F(y)}\right)^{j-2}(1-F(y))^{n}\frac{f(x)f(y)}{(1-F(y)^2}\,dx\,dy \\
&= \iint L_g^{(r)}(x-y)\left\{\left(\frac{1-F(x)}{1-F(y)}\right)^{n-1} - \left(\frac{F(y)-F(x)}{1-F(y)}\right)^{n-1}\right\}\frac{(1-F(y))^n}{(1-F(y))^2} \\
&\quad \times f(x)f(y)\,dx\,dy \\
&= \iint L_g^{(r)}(x-y)f(x)\lambda(y)\left\{(1-F(x))^{n-1} - (F(y)-F(x))^{n-1}\right\}\,dx\,dy.
\end{aligned}$$

For $i = 2$ we have

$$\begin{aligned}
\mathbb{E}\hat{\psi}_{r,2} &= \frac{1}{n-1}\iint \sum_{j=3}^{n} L_g^{(r)}(x-y)\frac{n(n-1)(n-2)!}{(j-3)!(n-j+1)!}F(x)(F(y)-F(x))^{j-3} \\
&\quad \times (1-F(y))^{n-j} f(x)f(y)\,dx\,dy \\
&= n\iint \sum_{j=3}^{n} L_g^{(r)}(x-y)\binom{n-2}{j-3}\left(\frac{F(y)-F(x)}{1-F(y)}\right)^{j-3}\frac{F(x)(1-F(y))^n}{(1-F(y))^3} \\
&\quad \times f(x)f(y)\,dx\,dy \\
&= n\iint L_g^{(r)}(x-y)F(x)\left\{\left(\frac{1-F(x)}{1-F(y)}\right)^{n-2} - \left(\frac{F(y)-F(x)}{1-F(y)}\right)^{n-2}\right\}(1-F(y))^n \\
&\quad \times \frac{f(x)f(y)}{(1-F(y))^3}\,dx\,dy \\
&= n\iint L_g^{(r)}(x-y)F(x)\left\{(1-F(x))^{n-2} - (F(y)-F(x))^{n-2}\right\}f(x)\lambda(y)\,dx\,dy.
\end{aligned}$$

Similarly for $i = 3$

$$\mathbb{E}\hat{\psi}_{r,3} = \frac{n(n-1)}{2!}\iint L_g^{(r)}(x-y)F(x)^2\left\{(1-F(x))^{n-3}-(F(y)-F(x))^{n-3}\right\} \times f(x)\lambda(y)\,dx\,dy$$

If we carry on like this we finally find,

$$\mathbb{E}\hat{\psi}_{r,n-1} = \iint L_g^{(r)}(x-y)F^{n-3}(x)\left\{(1-F(x))-(F(y)-F(x))\right\}f(x)\lambda(y)\,dx\,dy$$

$$\begin{aligned}\mathbb{E}\hat{\psi}_{r,n} &= \mathbb{E}\frac{L_g^{(r)}\left(X_{(n-1)}-X_{(n)}\right)w_{n-1,n}}{1\cdot 2} = \frac{1}{2}\iint L_g^{(r)}(x-y)\frac{n!}{(n-2)!}F^{n-2}(x)f(x)f(y)\,dx\,dy \\ &= \iint L_g^{(r)}(x-y)\frac{n!}{2(n-2)!}F^{n-2}(x)(1-F(y))f(x)\lambda(y)\,dx\,dy\end{aligned}$$

Using the fact that

$$\begin{gathered}(1-F(x))^{n-1}+nF(x)(1-F(x))^{n-2}+\frac{n(n-1)}{2}F^2(x)(1-F(x))^{n-3}+\dots \\ \dots+\frac{n(n-1)F^{n-2}(x)(1-F(y))}{2} = \sum_{i=1}^{n-1}\binom{n}{i-1}F^{i-1}(x)(1-F(x))^{n-i} = \\ \frac{(1-F(x))^n}{1-F(x)}\sum_{i=1}^{n-1}\binom{n}{i-1}\left(\frac{F(x)}{1-F(x)}\right)^{i-1} = \\ (1-F(x))^{n-1}\left(\frac{1}{(1-F(x))^n}-\frac{nF^{n-1}(x)}{(1-F(x))^{n-1}}-\frac{F^n(x)}{(1-F(x))^n}\right) = \\ \frac{1}{1-F(x)}-nF^{n-1}(x)-\frac{F^n(x)}{1-F(x)} \to \frac{1}{1-F(x)}\end{gathered}$$

as $n \to +\infty$, and that

$$\begin{gathered}(F(y)-F(x))^{n-1}+nF(x)(F(y)-F(x))^{n-2}+\frac{n(n-1)}{2}F^2(x)(F(y)-F(x))^{n-3}+\dots \\ +\frac{n(n-1)F^{n-1}(x)}{2} = \sum_{i=1}^{n-1}\binom{n}{i-1}F^{i-1}(x)(F(y)-F(x))^{n-i} = \\ \frac{(F(y)-F(x))^n}{F(y)-F(x)}\sum_{i=1}^{n-1}\binom{n}{i-1}\left(\frac{F(x)}{F(y)-F(x)}\right)^{i-1} = \\ (F(y)-F(x))^{n-1}\left\{\left(1+\frac{F(x)}{F(y)-F(x)}\right)^n-\frac{nF^{n-1}(x)}{(F(y)-F(x))^{n-1}}-\left(\frac{F(x)}{F(y)-F(x)}\right)^n\right\} = \\ (F(y)-F(x))^{n-1}\left\{\frac{F^n(y)}{(F(y)-F(x))^n}-\frac{nF^{n-1}(x)}{(F(y)-F(x))^{n-1}}-\frac{F^n(x)}{(F(y)-F(x))^n}\right\} = \\ \frac{F^n(y)}{F(y)-F(x)}-nF^{n-1}(x)-\frac{F^n(x)}{F(y)-F(x)} \to 0 \text{ as } n \to +\infty\end{gathered}$$

we find

$$\mathbb{E}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{L_g^{(r)}\left(X_{(i)}-X_{(j)}\right)w_{i,j}}{(n-j+1)(n-i+1)}=\iint L_g^{(r)}(x-y)f(x)\lambda(y)\times$$

$$\left\{\sum_{i=1}^{n-1}\binom{n}{i-1}F^{i-1}(x)(1-F(x))^{n-i}+\sum_{i=1}^{n-1}\binom{n}{i-1}F^{i-1}(x)(F(y)-F(x))^{n-i}\right\}dx\,dy$$
$$=\iint L_g^{(r)}(x-y)f(x)\lambda(y)\frac{1}{1-F(x)}\,dx\,dy=\iint L_g^{(r)}(x-y)\lambda(x)\lambda(y)\,dx\,dy.$$

Calculation of the means of $\hat{\psi}'_{r,i}$ is entirely similar with the only difference being that now integration takes place over the set $D_2=\{(x,y):y<x\}$. Hence,

$$\mathbb{E}\hat{\psi}_r=\iint_D L_g^{(r)}(x-y)\lambda(x)\lambda(y)\,dx\,dy$$

where $D=D_1+D_2$. ∎

Lemma A.1.2. *Let $X_1,\ldots,X_n$ be a sample from some density f having cdf F. Then, under the assumptions of theorem 2.5.1*

$$\mathbb{V}\mathrm{ar}\left(\hat{\psi}'_r(g)\right)=\iint L_g^{(r)}(x-y)^2\frac{\lambda(x)\lambda(y)}{(1-F(x))(1-F(y))}\,dx\,dy-\left(\mathbb{E}\hat{\psi}'_r\right)^2.$$

Proof.

$$\mathbb{V}\mathrm{ar}\left(\frac{1}{n^2}\sum\sum_{i\neq j}\frac{L_g^{(r)}\left(X_i-X_j\right)}{(1-F(X_i))(1-F(X_j))}\right)w_{i,j}=$$
$$\frac{1}{n^4}\sum\sum_{i\neq j}\mathbb{V}\mathrm{ar}\left\{\frac{L_g^{(r)}\left(X_i-X_j\right)w_{i,j}}{(1-F(X_i))(1-F(X_j))}\right\}+$$
$$\frac{1}{n^2(n-1)(n-2)}\sum\sum\sum_{i\neq j\neq k}\mathbb{C}\mathrm{ov}\left\{\frac{L_g^{(r)}\left(X_i-X_j\right)w_{i,j}}{(1-F(X_j)(1-F(X_i))},\frac{L_g^{(r)}\left(X_i-X_k\right)w_{i,k}}{(1-F(X_k))(1-F(X_i))}\right\}.$$

Now,

$$\mathbb{E}\frac{1}{n^4}\sum\sum_{i\neq j}\left\{\frac{L_g^{(r)}\left(X_i-X_j\right)}{(1-F(X_i))(1-F(X_j))}\right\}^2=\frac{n(n-1)}{n^4}\mathbb{E}\left\{\frac{L_g^{(r)}\left(X_1-X_2\right)}{(1-F(X_1))(1-F(X_2))}\right\}^2$$
$$=\left(\frac{1}{n^2}+o\left(n^{-2}\right)\right)\mathbb{E}\left\{\frac{L_g^{(r)}\left(X_1-X_2\right)}{(1-F(X_1))(1-F(X_2))}\right\}^2.$$

Also,

$$\mathbb{E}\left\{\frac{L_g^{(r)}\left(X_1-X_2\right)}{(1-F(X_1))(1-F(X_2))}\right\}^2=\iint L_g^{(r)}(x-y)^2\frac{\lambda(x)\lambda(y)}{(1-F(x))(1-F(y))}\,dx\,dy.$$

For the covariance, first note that

$$\sum\sum\sum_{i\neq j\neq k}\mathbb{C}\text{ov}\left\{\frac{L_g^{(r)}\left(X_i-X_j\right)}{(1-F(X_i))(1-F(X_j))},\frac{L_g^{(r)}\left(X_i-X_k\right)}{(1-F(X_k))(1-F(X_i))}\right\}=$$

$$4\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n}\mathbb{C}\text{ov}\left\{\frac{L_g^{(r)}\left(X_i-X_j\right)}{(1-F(X_i))(1-F(X_j))},\frac{L_g^{(r)}\left(X_i-X_k\right)}{(1-F(X_k))(1-F(X_i))}\right\}$$

and thus,

$$\frac{4}{n^2(n-1)(n-2)}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n}\mathbb{C}\text{ov}\left\{\frac{L_g^{(r)}\left(X_i-X_j\right)}{(1-F(X_i))(1-F(X_j))},\frac{L_g^{(r)}\left(X_i-X_k\right)}{(1-F(X_k))(1-F(X_i))}\right\}=$$

$$\frac{4n(n-1)(n-2)}{2n^2(n-1)(n-2)}\mathbb{C}\text{ov}\left\{\frac{L_g^{(r)}\left(X_1-X_2\right)}{(1-F(X_1))(1-F(X_2))},\frac{L_g^{(r)}\left(X_1-X_3\right)}{(1-F(X_3))(1-F(X_1))}\right\}.$$

But

$$\frac{2}{n}\mathbb{C}\text{ov}\left\{\frac{L_g^{(r)}\left(X_1-X_2\right)}{(1-F(X_1))(1-F(X_2))},\frac{L_g^{(r)}\left(X_1-X_3\right)}{(1-F(X_3))(1-F(X_1))}\right\}\leq$$

$$\frac{2}{n}\mathbb{E}\left\{\frac{L_g^{(r)}\left(X_1-X_2\right)}{(1-F(X_1))(1-F(X_2))}\frac{L_g^{(r)}\left(X_1-X_3\right)}{(1-F(X_3))(1-F(X_1))}\right\}.$$

Now,

$$\frac{2}{n}\mathbb{E}\left\{\frac{L_g^{(r)}\left(X_1-X_2\right)}{(1-F(X_1))(1-F(X_2))}\frac{L_g^{(r)}\left(X_1-X_3\right)}{(1-F(X_3))(1-F(X_1))}\right\}=$$

$$\frac{2}{n}\iiint\frac{L_g^{(r)}(x-y)}{(1-F(x))(1-F(y))}\frac{L_g^{(r)}(x-z)}{(1-F(x))(1-F(z))}f(x)f(y)f(z)\,dx\,dy\,dz=$$

$$\frac{2}{n}\iiint L_g(x-y)L_g(x-z)\frac{\lambda(x)}{1-F(x)}\lambda^{(r)}(y)\lambda^{(r)}(z)\,dx\,dy\,dz.$$

Applying the change of variables $y=x+gu$ and $z=x-gv$ and expanding in Taylor series around x,

$$\frac{2}{n}\mathbb{E}\left\{\frac{L_g^{(r)}\left(X_1-X_2\right)}{(1-F(X_1))(1-F(X_2))}\frac{L_g^{(r)}\left(X_1-X_3\right)}{(1-F(X_3))(1-F(X_1))}\right\}=$$

$$\frac{2}{n}\iiint L_g(u)L_g(v)\frac{\lambda(x)}{1-F(x)}\lambda^{(r)}(x+gu)\lambda^{(r)}(x-gv)\,dx\,du\,dv\simeq$$

$$\frac{2}{n}\int\lambda^{(r)}(x)^2\frac{\lambda(x)}{1-F(x)}\,dx\to 0\ \text{ as }\ n\to+\infty\qquad\blacksquare$$

A.2 Calculations, chapter 3.

A.2.1 Multiplication.

$$
\begin{gathered}
K(z)+\eta\left\{3zu'(0)K(z)+z^2u'(0)K'(z)\right\}+\\
\eta^2\left\{K(z)\left(3z^2u'(0)^2+\frac{3}{2}z^2u''(0)\right)+\frac{z^3}{2}u''(0)K'(z)+\frac{z^4}{2}u'(0)^2K''(z)+K'(z)3z^3u'(0)^2\right\}+\\
\eta^3\left\{K(z)z^3u'(0)^3+\frac{3}{3!}z^3u'''(0)K(z)+\frac{z^4}{3!}u'''(0)K'(z)+\frac{z^5}{2}u'(0)u''(0)K''(z)+\frac{z^6}{3}u'(0)^3K'''(z)\right.\\
\left.+\frac{3z^4}{2}u'(0)u''(0)K'(z)+\frac{3z^5}{2}u'(0)^3K''(z)+3z^4u'(0)^3K'(z)+\frac{3}{2}z^4u'(0)u''(0)K'(z)\right\}+\\
\eta^4\left\{\frac{z^5}{4!}u''''(0)K'(z)+\frac{z^6}{8}u''(0)^2K''(z)+\frac{z^6}{6}u'(0)u'''(0)K''(z)+\frac{1}{4}z^7u'(0)^2u''(0)K'''(z)+\right.\\
\frac{1}{4!}z^8u'(0)^4K''''(z)+z^4u'(0)u'''(0)K(z)+\frac{3}{2}z^4u'(0)^2u''(0)K(z)+\frac{3}{4}z^4u''(0)^2K(z)+\\
\frac{1}{8}z^4u''''K(z)+\frac{1}{2}z^5u'(0)u'''(0)K'(z)+\frac{3}{2}z^6u'(0)^2u''(0)K''(z)+\frac{1}{2}z^7u'(0)^4K'''(z)+\\
z^5u'(0)^4K'(z)+\frac{1}{2}z^5u'(0)u'''(0)K'(z)+3z^5u'(0)^2u''(0)+\frac{3}{2}z^5u'(0)^2u''(0)K'(z)+\\
\left.\frac{3}{2}z^6u'(0)^4K''(z)+\frac{3}{4}z^5u''(0)^2K'(z)+\frac{3}{4}z^6u''(0)u'(0)^2K''(z)\right\}.
\end{gathered}
$$

Taking common factors the product becomes:

$$
\begin{gathered}
K(z)+\eta\left\{3zu'(0)K(z)+z^2u'(0)K'(z)\right\}+\\
\eta^2\left\{K(z)\left(3z^2u'(0)^2+\frac{3}{2}z^2u''(0)\right)+K'(z)\left(\frac{z^3}{2}u''(0)+3z^3u'(0)^2\right)+\frac{z^4}{2}u'(0)^2K''(z)\right\}+\\
\eta^3\left\{K(z)\left(z^3u'(0)^3+\frac{3}{3!}z^3u'''(0)\right)+K'(z)\left(\frac{z^4}{3!}u'''(0)+\frac{3z^4}{2}u'(0)u''(0)+3z^4u'(0)^3+\right.\right.\\
\left.\left.\frac{3}{2}z^4u'(0)u'''(0)\right)+K''(z)\left(\frac{z^5}{2}u'(0)u''(0)+\frac{3z^5}{2}u'(0)^3\right)+K'''(z)\frac{z^6}{3}u'(0)^3\right\}+\\
\eta^4\left\{u''(0)\left(\frac{3}{4}z^4K(z)\right)+u'(0)^2u''(0)\left(\frac{3}{2}z^4K(z)+\frac{9}{2}z^5K'(z)+\frac{9}{4}z^6K''(z)+\frac{1}{4}z^7K'''(z)\right)+\right.\\
u''(0)^2\left(\frac{z^6K''(z)}{8}+\frac{3}{4}z^4K(z)+\frac{3}{4}z^5K'(z)\right)+u'(0)u'''(0)\left(z^4K(z)+z^5K'(z)+\frac{z^6K''(z)}{6}\right)+\\
\left.u'(0)^4\left(\frac{1}{4!}z^8K''''(z)\frac{3}{2}z^6K''(z)+\frac{1}{2}z^7K'''(z)+z^5K'(z)\right)+u''''(0)\left(\frac{1}{8}z^4K(z)+\frac{z^5}{4!}K'(z)\right)\right\}.
\end{gathered}
\tag{A.6}
$$

REFERENCES

[1] I. Abramson. Arbitrariness of the Pilot Estimator in Adaptive Kernel Methods. *Journal of Multivariate Analysis*, 12:562–567, 1982.

[2] I. Abramson. On Bandwidth Variation in Kernel Estimates-a square root law. *The Annals of Statistics*, 27(3):1217–1223, 1982.

[3] R. Barlow and F. Proschan. *Statistical Theory of Reliability and Life Testing Probability Models.* Holt, Rinehart and Winston Inc., 1975.

[4] M.S. Bartlett. Statistical Estimation of Density Functions . *Sankhya, Ser. A*, 25:245–254, 1963.

[5] W. Breiman, L. Meisel and E. Purcell. Variable kernel estimates of probability densities. *Technometrics*, 19:135–144, 1977.

[6] D.L. Burkholder. Distribution Function Inequalities for Martingales. *The Annals of Probability*, 1, (1):19–42, 1973.

[7] M. Y. Cheng. On Boundary Effects of Smooth Curve Estimates. *PhD Dissertation, Univ. of North Carolina at Chapel Hill.*, 1994.

[8] W.S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *J. of the American Statistical Association*, 74:829–835, 1979.

[9] D.R. Cox and D. Oakes. *Analysis of Survival Data* . Chapman and Hall, 1984.

[10] L. Devroye and L. Gyorfi. *Nonparametric Density Estimation, The L_1 View*. Wiley, 1984.

[11] R.C. Elandt-Johnson and N.L. Johnson. *Survival Models and Data Analysis*. Wiley, 1980.

[12] J. Fan. Design Adaptive Nonparametric Regression. *Journal of the American Statistical Association*, 87:998–1004, 1992.

[13] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, 1997.

[14] P. Hall. On the Bias of Variable Bandwidth Curve Estimators. *Biometrika*, 77, (3):529–35, 1990.

[15] P. Hall and C. Heyde. *Martingale Limit Theory and its Applications.* Academic Press, New York, 1980.

[16] P. Hall and J.S Marron. Variable Window Width Kernel Estimates of Probability Densities. *Probability Theory and Related fields*, 80:37–49, 1988.

[17] P. Hall, T. Hu and J.S. Marron. Improved variable window kernel estimates of probability densities. *The Annals of Statistics*, 23, (1):1–10, 1995.

[18] J.S. Jones, M.C. Marron and S.J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91, 433:401–407, 1996.

[19] M.C. Jones. Variable Kernel Density Estimates and Variable Kernel Density Estimates. *Australian Journal of Statistics*, 32:361–371, 1990.

[20] J.D. Kalbfleisch and R.L. Prentice. *The Statistical Analysis of Failure Time Data.* Wiley, 1980.

[21] J.F. Lawless. *Statistical Models and Methods for Lifetime Data.* Wiley, 1982.

[22] S. Lo, Y. Mack and J. Wang. Density and hazard rate estimation for censored data via strong representation of the kaplan-meyer estimator. *Probability Theory and Related Fields*, 80:461–473, 1989.

[23] Y. Mack and M. Rosenblatt. Multivariate k-nearest Neighbor Density Estimates. *J. of Multivariate Analysis*, 9:1–15, 1979.

[24] J.S. Marron and M.P. Wand. Exact Mean Integrated Square Error. *The Annals of Statistics*, 20(2):712–736, 1992.

[25] S.K. McCune and E.D. McCune. On improving convergence rates for nonnegative kernel failure-rate function estimators. *Statist. Probability Letters*, 6(2):71–76, 1987.

[26] B. Park and J.S. Marron. Comparison of Data Driven Bandwidth Selectors. *Journal of the American Statistical Association*, 85(409):66–72, 1962.

[27] E. Parzen. On Estimation of a Probability Density Function and Mode . *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[28] P.N. Patil. Bandwidth Choice for Non Parametric Hazard Rate Estimation . *Journal of Statistical Planning and Inference*, 35:15–30, 1993.

[29] P.N. Patil. On The Least Squares Cross Validation Bandwidth In Hazard Rate Estimation. *The Annals of statistics*, 21:1792–1810, 1993.

[30] P.N. Patil, M. Wells and J.S. Marron. Some heuristics of kernels based estimators of ratio functions. *Nonparametric Statistics*, 4:203–209, 1994.

[31] V. Petrov. *Limit Theorems of Probability Theory*. Oxford Science Publications, 1995.

[32] D. Pollard. *Convergence of Stochastic Processes*. Springer, Berlin, 1983.

[33] P. Rao. *Non Parametric Functional Estimation*. Academic Press, New York, 1983.

[34] J. Rice and M. Rosenblatt. Estimation of the Log-Survivor Function and Hazard Function . *Sankya Ser. A*, 38:60–78, 1976.

[35] D. Ruppert and D. Cline. Bias Reduction in Kernel Density Estimation by Smoothed Empirical Transformations. *Annals of Statistics*, 22, (1):185–210, 1994.

[36] P. Sarda and P. Vieu. Smoothing Parameter Selection In Hazard Rate Estimation. *Statistics and Probability Letters*, 11:429–434, 1991.

[37] R. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley and Sons, 1980.

[38] S.J. Sheather and M.C. Jones. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *J. of the Royal Statistical Society, B*, 53(3):683–690, 1991.

[39] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 1986.

[40] J. Simonoff. *Smoothing Methods in statistics*. Springer, 1996.

[41] N. Singpurwalla and M. Wong. Improvement of Kernel Estimators of the Failure Rate Function Using the Generalized Jacknife. *Nonparametric Statistics*, 4:203–209, 1994.

[42] C.J. Stone. Consistent Nonparametric Regression. *Annals of Statistics*, 5, (4):595–645, 1977.

[43] C.J. Stone. An Asymptotically Optimal Window Selection Rule For Kernel Density Estimates. *Annals of Statistics*, 12, (4):1285–1297, 1984.

[44] M. Tanner and W. Wong. The Estimation of the Hazard Function from Randomly Censored Data by the Kernel Method. *The Annals of Statistics*, 11(3):989–993, 1983.

[45] G.R. Terrel and D.W. Scott. Variable Kernel Density Estimation. *The Annals of Statistics*, 20:1236–1265, 1992.

[46] L. Tierney. *Lisp-Stat: An object oriented enviroment for statistical computing and dynamic graphics* . Wiley and Sons, 1990.

[47] M. Wand and M.C. Jones. *Kernel Smoothing*. Chapman and Hall, 1995.

[48] G. Watson and M. Leadbetter. Hazard Analysis I. *Biometrika*, 51(1 and 2):175–184, 1964.

[49] G. Watson and M. Leadbetter. Hazard Analysis II. *Sankhya, Ser. A*, 26:101–116, 1964.